Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department of Forest Mycology and Plant Pathology

Development of expression markers for reduced susceptibility in *Fraxinus excelsior*

Matilda Stein Åslund

Independent project • 15 credits
Biology and Environmental Sciences
Uppsala 2018

Development of expression markers for reduced susceptibility in *Fraxinus excelsior*

Matilda Stein Åslund

Supervisor: Malin Elfstrand, Swedish University of Agricultural Sciences, Department of Forest Mycology and Plant Pathology
Assistant supervisor: Audrius Menkis, Swedish University of Agricultural Sciences, Department of Forest Mycology and Plant Pathology
Assistant supervisor: Heriberto Vélëz, Swedish University of Agricultural Sciences, Department of Forest Mycology and Plant Pathology
Examiner: Mukesh Dubey, Swedish University of Agricultural Sciences, Department of Forest Mycology and Plant Pathology

Credits: 15 credits
Level: First cycle, G2E
Course title: Independent project in Biology
Course code: EX0689
Programme/education: Biology and Environmental Sciences

Place of publication: Uppsala
Year of publication: 2018
Online publication: https://stud.epsilon.slu.se

Keywords: *Fraxinus excelsior*, European ash, common ash, molecular markers, expression markers, ash dieback disease, *Hymenoschyphus fraxineus*

Swedish University of Agricultural Sciences
Faculty of Natural Resources and Agricultural Sciences
Department of Forest Mycology and Plant Pathology

Abstract

Common ash (*Fraxinus excelsior*) extends throughout temperate Europe, and in southern and central Sweden it is one of the most common broad-leaved deciduous trees. It is economically valuable thanks to its elastic, hard and pressure-resistant wood, and pollarded ash trees provide a cultural heritage in many European landscapes. The species is also important for biodiversity, as many organisms depend on it. *F. excelsior* is currently suffering from ash dieback disease, an epidemic caused by the fungus *Hymenoscyphus fraxineus*. Ash dieback lethally affects trees of all ages. Common ash is listed as 'endangered' in the Red list of Sweden since 2015, due to trees being affected by dieback all over the *F. excelsior* distribution range in the country.

Molecular markers for resistance or tolerance in plants have been studied mainly with the aim to improve plant breeding in agriculture. Using tolerance-predictive markers, trees with reduced susceptibility to pathogens can be identified. To identify markers, score them and correlate the variation in both gene sequence and expression with variation in traits associative transcriptomics are of use. Using this method in a study on F. excelsior, Harper et al. (2016) identified a SNP as a predictor of reduced susceptibility to ash dieback. In another study on common ash and dieback by Sahraei (2016), 1082 differentially expressed genes in susceptible and tolerant clones were revealed. In this study, the aim was to identify and evaluate molecular markers for reduced susceptibility in F. excelsior. Phenotypically classified tolerant and susceptible ash trees located in Gotland, Sweden, were used. To evaluate the SNP, Sanger sequencing was used to score the presence of the polymorphism. To identify molecular markers for reduced susceptibility, ten candidate genes out of the 1082 differentially expressed genes were tested as possible markers for tolerance using qPCR. The SNP was shown to be a moderately good predictor of lower susceptibility to dieback. One of the genes in the study showed a significant difference in relative expression between tolerant and susceptible individuals (p-value <0.05).

Keywords: Fraxinus excelsior, common ash, European ash, ash dieback disease, *Hymenoscyphus fraxineus*, expression markers, molecular markers

Sammanfattning

Asken (*Fraxinus excelsior*) har det tempererade Europa som utbredningsområde, och i södra och centrala Sverige är den ett av de vanligaste bredbladiga lövträden. Asken är ekonomiskt värdefull tack vare det elastiska, hårda och tryckbeständiga träet, och hamlade askar utgör ett kulturellt arv i många europeiska landskap. Arten är också viktig för biologisk mångfald, eftersom många organismer är beroende av den. Askpopulationen är för närvarande starkt påverkad av askskottsjukan, en epidemi orsakad av svampen *Hymenoscyphus fraxineus*. Askskottsjukan påverkar träd i alla åldrar, ofta med dödlig utgång. Rödlistningsbedömningen i Sverige är sedan 2015 ”starkt hotad”, på grund av att askträd påverkas av askskottsjukan i hela trädets utbredningsområde i landet.

Molekylära markörer för resistens eller tolerans i växter har studerats huvudsakligen i syfte att förbättra växtförädling inom jordbruket. Med hjälp av tolerans-prediktiva markörer kan träd med nedsatt mottaglighet för patogener identifieras. För att identifiera markörer, värdera dem och korrelera variationen i både gensekvens och uttryck med variation i egenskaper är *associative transcriptomics* användbart. Med hjälp av denna metod i en studie på *F. excelsior*, identifierade Harper *et al*. (2016) en SNP som markör för minskad mottaglighet för askskottsjukan. I en annan studie på ask och askskottsjukan, av Sahraei (2016), visade sig 1082 gener vara differentiellt uttryckta i mottagliga och toleranta kloner. I denna studie var syftet att identifiera och utvärdera molekylära markörer för minskad patogenmottaglighet hos *F. excelsior*. Fenotypiskt klassificerade toleranta och mottagliga askar på Gotland användes. För att utvärdera den SNP som tidigare identifierats användes Sanger-sekvensering för analys av polymorfismen. För att identifiera molekylära markörer för minskad patogenmottaglighet användes qPCR, och tio kandidatgener av de 1082 differentiellt uttryckta generna analyserades. Harpers SNP visade sig vara en måttligt bra markör för lägre mottaglighet för askskottsjukan. En av generna i studien visade en signifikant skillnad i relativt uttryck mellan toleranta och mottagliga individer (p-värde <0,05).

Table of contents

Abbreviations

CAZY	carbohydrate-active enzyme
cDNA	complementary DNA
cSNP	cDNA-based single-nucleotide polymorphism
GEM	gene expression marker
LRR	leucine rich repeat
qPCR	quantitative polymerase chain reaction
PCR	polymerase chain reaction
PR	pathogenesis related protein
TF	transcription factor

1 Introduction

1.1 *Fraxinus excelsior*

Common ash (*Fraxinus excelsior)* extends throughout temperate Europe (More, 2013). In Sweden, its northern boundary follows *Limes Norrlandicus*, and in southern and central Sweden *F. excelsior* is one of the most common broad-leaved deciduous trees (Artfakta, 2015). It is potentially favoured by climate change, since it is frost sensitive (Pautasso *et al.*, 2013) and drought tolerant (Scherrer, Bader and Körner, 2011). Ash acts as a pioneer species in woodland recolonisation, and a general expansion of the species has come as a result of its ecological and physiological flexibility and occurring rural abandonment (Marigo *et al.*, 2000). In the past, the species was the most economically valuable tree in Europe (More, 2013). The noble hardwood is elastic, hard and resistant to pressure, and is therefore in high demand for production of furniture, veneer, composite wood, flooring, tool handles and sport equipment (Pautasso *et al*., 2013 and sources within). Pollarded ash trees are iconic elements of landscapes throughout Europe, providing a cultural heritage from times when the leaves supplied fodder for livestock during drought. To rare epiphytic lichens, the pollarded trees contribute with invaluable habitat (Pautasso *et al*., 2013 and sources within). Common ash has a keystone role in European floodplain forest ecosystems. There is little knowledge of how much biodiversity relies on ash trees, but many organisms depend on them (e.g., wood decaying fungi, saproxylic insects, epiphytic lichens, geophytes and birds). 308 fungi are listed as occurring on *F. excelsior*. Old ash trees are particularly important regarding biodiversity conservation, but young trees are also valuable for some threatened species. To land mollusc diversity, common ash is a keystone species (Pautasso *et al*., 2013 and sources within). Common ash is crucial to the Swedish biodiversity and many threatened species are

associated with the tree, for example, the jewel beetle *Agrilus convexicollis*, the lichen *Pyrenula nitidella*, the poroid fungus *Vanderbylia fraxinea*, and the butterfly *Euphydryas maturna* (Artfakta, 2015).

1.2 Ash dieback disease

Common ash is currently suffering from an epidemic disease called ash dieback (Cleary *et al.*, 2014). The disease is caused by the ascomycete fungus *Hymenoscyphus fraxineus* (Stenlid *et al.*, 2017), which lethally affects trees of all ages (Pautasso *et al.*, 2013). Ash dieback was first observed in Poland in the beginning of the 1990s (Pautasso *et al.*, 2013), and by 2009, it had spread to 13 European countries (Johanson, 2009). *H. fraxineus* is an endophyte or weakly pathogenic fungus in its region of origin in Asia (Cleary *et al.*, 2016). The life cycle of the fungus involves spore infection on healthy leaves, or sometimes bark tissue, during summer (Harper *et al.*, 2016; Stenlid *et al.*, 2017), which is followed by growth along the petiole into twigs (Gross *et al.*, 2014). After defoliation the fungus remains in the petioles and releases wind-borne spores the following summer (Harper *et al.*, 2016). Unrestrained growth in inner bark and xylem results in dieback of the crown (Stenlid *et al.*, 2017), as it prevents the supply of water and nutrients (Johanson, 2009). This can be fatal for the host, and mortality levels are high (Pautasso *et al.*, 2013). Early symptoms of the disease are brown necrosis on leaves, leaf nerves and leaf scars, dry buds and dead shoots. Wilted leaves are common during the summers. Cankers on branches and stem as well as dead branches and shoots are signs of a wide-spread infection. New shoots replace dead top shoots due to apical dominance and epicormic shoots succeed dead branches, and the mixture of dead, infected and new twigs give the tree a bushy appearance (Johanson, 2009). *F. excelsior* has been listed as 'endangered' in the Red list of Sweden since 2015, due to ash trees being affected by dieback all over the species distribution range in the country (Artfakta, 2015).

1.2.1 Ash dieback disease on Gotland

Gotland is an island which is unique regarding its richness in deciduous trees. It has a bedrock of limestone and therefore a relatively high pH, which benefits *F. excelsior*. The pathogen pressure from *H. fraxineus* on Gotland is high, and a large proportion of the common ash trees have dieback-related symptoms (Audrius Menkis, personal conversation). Gotland is therefore a well-suited location for studies on ash dieback and thus the sample material in this study comes from this island.

1.3 Disease and resistance/tolerance to pathogens

In plant pathology, disease is defined as the malfunctioning of the host cells and tissues as a result of the continuous irritation by pathogens or environmental factors. A pathogen's ability to damage or infect a host is defined as virulence (Agrios, 2005).

Resistance traits are defined as traits that are reducing host contact with pathogens and, once the host has been infected, they reduce pathogen growth rate. The definition of tolerance to pathogen infection is the host's ability of reducing the infection effect on fitness (Kover and Schaal, 2002).

The studies reviewed for this project use both concepts of resistance and tolerance. As resistance is difficult to prove, and as I consider the error caused by misusing the expression of resistant for tolerant organisms is more serious than the opposite, the concept of tolerance is used in this study.

1.3.1 Molecular markers

A molecular marker is a genetic polymorphism or molecular component that is correlated with a trait of interest. Markers can be used as experimental probes to keep track of individuals, tissue, cells or genes. In mapping, molecular markers are used to fill the chromosomal intervals between genes of known phenotypes, providing a higher resolution than previous methods as they were limited to genes with variant alleles encoding observable phenotypes (Griffiths *et al.*, 2000).

Molecular markers for resistance or tolerance in plants have been studied mainly with the aim to improve plant breeding in agriculture. Even though statistical analyses in some cases have revealed highly significant results, mismatches between expected phenotype and marker expression are common (Lopez-Pardo *et al.*, 2013). The literature presents varying results regarding the coincidence between expression of a marker and resistance. Markers that separate susceptible and resistant phenotypes perfectly and are selected as diagnostic tools to detect resistance in one study, might show disagreement between phenotype and marker analysis in others. Many markers are considered tools for detecting resistant clones if only a few of the resistant genotypes are susceptible. However, despite showing susceptibility, in some cases they are actually resistant. Mismatches can be due to recombination between marker and the resistance gene, individuals escaping inoculation, errors in PCR assays or ELISA (enzyme-linked immunosorbent assay) or effects of timing of the evaluation. Several identified markers for resistance are disproved as reliable

in later studies, and the misclassifications can be explained by natural allelic variation with a lower efficiency, that might only provide full resistance with additional genes. Usually, the resistance markers are in linkage disequilibrium with causal genes, but are not expected to actually cause the phenotype. Breeding with material wrongly classified as resistant might be worse than to discard genotypes incorrectly classified as susceptible (Lopez-Pardo *et al.*, 2013).

Using tolerance-predictive molecular markers, trees with reduced susceptibility to pathogens can be identified. Their genotypes will be broadly different, in addition to at the predictive marker. Using these trees for reforestation increases the chances of sustaining a genetically diverse population, compared to strategies where breeding from a few, highly tolerant individuals is done. The latter strategy decreases the genetic diversity and, thereby, increases the risk of the population being heavily affected by the next disease or pest arriving (Harper *et al.*, 2016).

1.3.2 cSNP indicating reduced susceptibility to ash dieback

Association genetics can describe regions of the genome where traits are controlled, and provide markers to facilitate marker-assisted breeding. Since most plants are polyploid it is difficult to identify the desired markers, and ordering these markers by assembling the genome sequence is complicated. Associative transcriptomics uses sequencing of the transcriptome to identify markers, score them and correlate the variation in both gene sequence and expression with variation in traits (Harper *et al.*, 2012).

Harper *et al.* (2016) analysed 182 common ash trees in Denmark using associative transcriptomics in a plant pathology study, where variants in gene sequence and gene expression that scored for symptoms of disease were discovered. Markers associated with canopy damage in trees infected by H. fraxineus were identified and used to predict phenotypes in a test panel of trees, and thereby they could identify individuals with a low level of susceptibility to dieback. The leaves from which the markers were detected were uninfected, suggesting the mechanism is different from pathogen induced resistance (Harper *et al.*, 2016).

In the study, a cDNA-based SNP (or cSNP), called Gene_22343_Predicted_mRNA_scaffold3139:2378, was identified as a moderately good predictor (p-value <0.01) of reduced susceptibility to dieback. The cSNP is, based on its BLAST hits, a member of the MADS box transcription factor family. The study revealed that several gene models with the best associations with susceptibility to diseases are part of the MADS box transcription factor family as well, indicating

markers within a regulatory network of genes can be associated with reduced susceptibility to ash dieback disease (Harper *et al.*, 2016). The polymorphism associated with reduced disease susceptibility consists of two alleles, both A and G bases, while the other is an A base alone as shown in Figure 1. Homozygosity for the G allele does not seem to be present in the studied *F. excelsior* material.

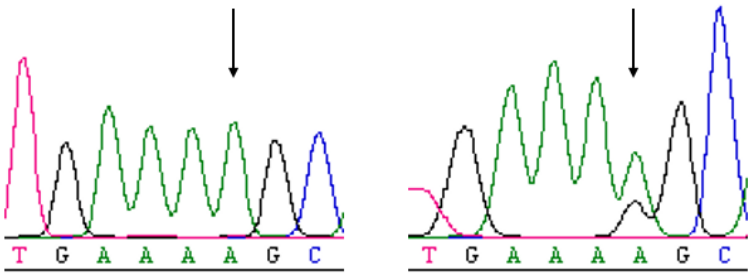


Figure 1. The cSNP position, indicated by arrows, showed in sequencing chromatograms. The G base is either absent (left, susceptible), or present at a low level compared to the A base (right, tolerant).

Sanger sequencing

Sanger sequencing is used to analyse DNA sequences. The sequencing is done by synthesis of DNA chains randomly terminated by added dideoxynucleotides (ddNTPs), followed by analysis of the fragment lengths to determine the DNA sequence (Klug *et al.*, 2014). The Sanger sequencing method originally utilised four 2′, 3′-ddNTPs in the polymerase, one for each nucleotide. ddNTPs, unlike dNTPs, lack the 3′-hydroxyl group. The ddNTPs terminate further elongation of the DNA chain, since the 3′-OH is required for the phosphodiester bond between the nucleotide and primer terminus. Today, fluorescently labeled dye-terminators (chain-terminating ddNTPs) are utilised. A distinct fluorophore is attached to each ddNTP, letting the chain termination reactions happen in one single reaction instead of four separate ones (Chen, 2014).

1.3.3 Gene expression related to tolerance

A previous study on *F. excelsior* shows that 1082 genes are differentially expressed in susceptible and resistant clones (Sahraei, 2016). The phenotypic classifications were based on the relationship between genetic values (BLUPs, or Best Linear Unbiased Predictions), such as vitality and damage traits, and dieback-related damage scores. The damage was scored from 0 (none) to 9 (very serious damage) for the entire crown by Stener (2018).

Gene expression profiling with qPCR

qPCR, or quantitative real-time PCR, uses fluorescent probes or dyes to measure the DNA or RNA product quantity present after each cycle of amplification (Klug *et al.*, 2014). The cycle threshold, C_T, reports when fluorescence levels becomes detectable over a defined threshold, and the cycle number at which this happens is proportional to number of copies in the template (Berg *et al.*, 2015).

1.4 Objectives

The aim of this study was to identify and evaluate molecular markers for reduced susceptibility in *F. excelsior* with the guidance of existing data for differentially expressed genes in relation to tolerance to *H. fraxineus*.

2 Materials and Methods

2.1 Validation of existing molecular marker

2.1.1 Sample collection

Leaf tissue was collected from 60 common ash trees (*Fraxinus excelsior*) in Gotland, Sweden, right after time of flushing in June 2017. Sampling was made from 50 tolerant and ten susceptible individuals based on phenotyping made in autumn 2013, and the classification as tolerant or susceptible was based on the occurrence of dieback-related symptoms (Audrius Menkis, personal conversation). Leaf tissue was collected in Falcon tubes with RNA*later* RNA Stabilization Reagent (Qiagen) and stored at -20°C.

2.1.2 RNA extractions

I milled approximately 2-4 cm^2 leaf tissue from each individual using a mortar and pestle, which had been baked at 550°C for 2-4 hours in order to ensure RNases, that degrade the RNA, were eliminated. The leaf tissue was soaked in liquid nitrogen during milling. The powdered tissue and liquid nitrogen were poured into Eppendorf tubes and incubated in a water bath (65°C) for 2 minutes to lyse cell walls and protein complexes, and RNA was thereafter extracted using the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's protocol, where β-mercaptoethanol and Buffer RLT (Qiagen) were added. I quantified the amount of nucleic acids in the extractions using NanoDrop spectrophotometer. Samples with low concentrations of RNA were re-extracted but excluded from the experiment if the extraction failed a second time. The extracted RNA was stored at -80°C.

To remove DNA contamination, I treated 2 µg RNA from each sample with DNase I (Sigma-Aldrich) as stated by the manufacturer. I analysed the quality of the DNased RNA-extractions using Agilent 2100 bioanalyzer (Agilent Technologies) in accordance with the manufacturer's protocol, and samples with an RIN-number (RNA integrity number) >4.5 were considered adequate for continued analysis. Samples with an RIN-number <4.5 were re-extracted once, but excluded if the second quality analysis failed.

2.1.3 cDNA synthesis

Complementary DNA (cDNA) was used to exclude introns in the amplification. cDNA synthesis was performed on 1 µg total RNA using iScript cDNA Synthesis Kit (Bio-Rad Laboratories) as per instructions from the manufacturer with the modification of 90 minutes at 42°C instead of 30 minutes. I made this adjustment to ensure synthesis was fully executed, as the reaction can be slowed down due to non-optimal reaction conditions caused by impurities in the leaf tissue. The cDNA was stored at -20°C.

2.1.4 Primers

To score the cSNP Gene_22343_ Predicted_mRNA_scaffold3139:2378 the primers described by Harper *et al.* (2016) were used, see Appendix 1. The primers were ordered from Eurofins Genomics (Ebersberg, Germany).

2.1.5 PCR

I amplified the cSNP by PCR, with a master mix for each reaction with the following composition: 15.0 µL DreamTaq Green PCR Master Mix (2X) (Thermo Fisher Scientific), 12.8 µL nuclease-free water, 0.6 µL forward and reverse primer (see Appendix 1), and 1.0 µL cDNA template. The PCR was programmed according to Sollars *et al.* (2017); 1 cycle of 5 min at 94°C, 15 cycles of 30 sec at 94°C, 30 sec at 63°C with a decreasing temperature of -1°C per cycle and 1 min at 72°C, 30 cycles of 30 sec at 94°C, 30 sec at 53°C and 1 min at 72°C followed by 7 min at 72°C. The touch-down programme was used to ensure optimal amplification, since a higher temperature gives a higher specificity, but might give less product.

2.1.6 Gel electrophoresis and DNA extraction from gel

The PCR product was analysed using gel electrophoresis. I used TAE buffer in a 1.0% agarose gel with Nancy-520 (Sigma-Aldrich) as fluorescent stain. GeneRuler

DNA ladder (Thermo Fisher Scientific) was used for DNA sizing and the electrophoresis was run for 80 minutes at 120 V. The gel was thereafter examined in UV-light. Due to uncertainty regarding primer dimers and degraded DNA, I ran another gel electrophoresis with three of the samples, excised all bands using a scalpel, purified the DNA as described below, and sent them to Macrogen (Amsterdam, Netherlands) for Sanger sequencing to verify the target product. When the fragment length of the product was confirmed (see Figure 2), I ran a second gel electrophoresis with the samples that had amplicons in the first run, and excised the desired band.

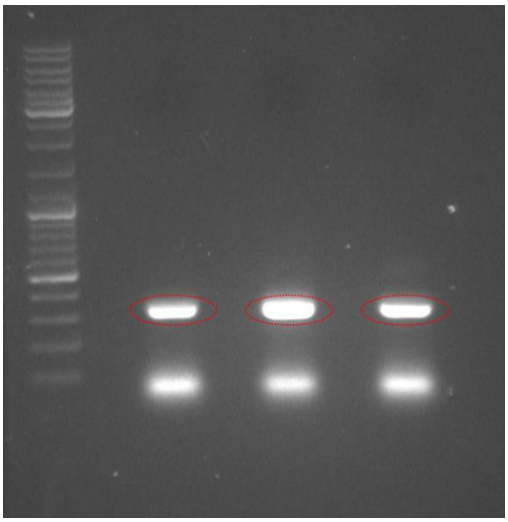


Figure 2. Gel image of cSNP amplicons for verification of target product. Circles indicate correct PCR product.

The excised DNA fragment bands were purified using GeneJET Gel Extraction Kit (Thermo Fisher Scientific) as specified by the manufacturer, with the modifications of added isopropanol and an additional washing step. Isopropanol was added to the reaction and Binding Buffer solution as it facilitates the precipitation of DNA, and the additional washing was done twice to ensure all unwanted content was removed. The purified DNA was stored at -20°C.

2.1.7 Sequencing and analysis

Purified DNA from 40 tolerant and eight susceptible ash tree individuals were sent to Macrogen (Amsterdam, Netherlands) for Sanger sequencing. I analysed the sequencing trace files using SeqMan Pro (DNASTAR). After aligning all the sequences using the default parameters of the software, I scored the cSNP (Gene_22343_ Predicted_mRNA_scaffold3139:2378) based on the presence of one peak for A (i.e., one allele), or the presence of a double-peak consisting of A and G

(i.e., both alleles). I then compared the data from the genotype analysis to the phenotype analysis that the sample collection was based on.

2.2 Relative expression of candidate defence genes

2.2.1 Gene expression data

I was provided with gene expression data from four healthy *F. excelsior* clones of which two were tolerant and two were susceptible to *H. fraxineus*, classified based on dieback damage scores as described in section 1.3.3.

The data was derived by extracting RNA from bark tissue for cDNA synthesis, and the cDNA samples were sent for sequencing on an Illumina HiSeq2500 at SciLife-Lab (Stockholm, Sweden) as described in a previous study (Sahraei, 2016).

2.2.2 Selection of candidate genes

Candidate genes were selected from the 1082 differentially expressed genes from the experiment presented in section 1.3.3. Genes were chosen based on expression data (number of reads >0 in both tolerant and susceptible individuals), size (>300 bp), and BlastX hits (>0). The genes that lacked BlastX hits were considered to have a fragment size too short for this analysis, and are possibly not ash genes. A similarity mean >0.75 for the BlastX hits was required. The remaining genes were evaluated and 15 genes were selected based on gene ontology annotations, with the aim of successful analysis of approximately ten genes.

2.2.3 Sample selection

I selected 19 samples of cDNA from the Gotland population of common ash trees (used in the cSNP analysis in section 2.1) based on the phenotype and genotype combined, with the purpose of analysing individuals classified as tolerant against ash dieback disease by both Menkis and Harper (MTHT), susceptible according to both Menkis and Harper (MSHt), tolerant according to Menkis but susceptible according to Harper (MTHt), and samples classified as susceptible by Menkis but tolerant according to Harper (MSHT).

2.2.4 Primers

I used coding sequences for the selected genes to design primers using Primer3 (Untergasser *et al.*, 2018). Primer3 was set to design primers with a size product range of 75-150 bp (optimal 75-120 bp), primer size 18-25 bp (optimal 20 bp), annealing temperature 60-63°C (optimal 60°C) with a maximum temperature difference of 2°C between forward and reverse primer, a primer nucleotide G or C content of 40-60%, maximum self-complementary of 5 bp and maximum self-complementary 3' end of 3 bp. From the primer suggestions I chose primers avoiding palindromes and excessive numbers of G/C or A/T nucleotides in succession, especially in the 3' ends. Palindromes might fold double, form hairpins, create background amplification or bind poorly. The primers were ordered from Eurofins Genomics (Ebergsberg, Germany).

Primer testing - PCR

To test the primers, pooled cDNA from the selected samples was used in PCR. The master mix composition for each reaction was 1.0 µL (10 mM) forward and reverse primer (see Appendix 1), 12.8 µL nuclease-free water, 2.0 µL DreamTaq Green Buffer (Thermo Fisher Scientific), 2.0 µL dNTP (deoxynucleotide), 0.2 µL DreamTaq DNA polymerase (Thermo Fisher Scientific) and 1.0 µL cDNA template. The PCR cycling conditions were: 5 min at 95°C, 35 cycles of 30 sec at 95°C, 30 sec at 57°C and 30 sec at 72°C, followed by 7 min at 72°C.

Primer testing - gel electrophoresis and PCR purification

The PCR product for each primer pair was analysed and purified as described in section 2.1.6, with the adjustments of a higher agarose percentage in the gel (1.5%) and an electrophoresis run set for 100 minutes at 100 V. I examined the gel in UV-light to ensure amplification of the desirable DNA fragments.

I quantified the purified PCR product using Invitrogen Qubit 4 Fluorometer (Thermo Fisher Scientific) following the manufacturer's protocol.

Standards for candidate genes

Based on the concentration of purified PCR products, dilutions from stock concentrations of 10^8 or 10^9 copies per µL for each candidate gene was calculated using the Thermo Fisher DNA Copy Number and Dilution Calculator (Thermo Fisher Scientific). Serial dilutions were made down to 10^2 copies per µL.

The standards were tested using qPCR. The master mix composition was 7.5 µL Sso Fast EvaGreen supermix (Bio-Rad Laboratories), 5.0 µL nuclease-free water and

0.75 µL (10 mM) forward and reverse primer. I loaded 14.0 µL of the master mix in each well of the PCR plate and then 1.0 µL standard dilutions from 107 to 103 copies per µL. Reactions were setup in triplicate and included negative controls for each master mix. The qPCR was programmed for 30 sec at 95°C, 40 cycles of 5 sec at 95°C and 10 sec at 60°C, followed by 61 cycles of 11 sec at 65°C with a temperature change of +0.5°C per cycle. The primer pairs that amplified the targeted product, gave a clear melting curve, and had an efficiency value of 80-120% were selected for the experiment. Ten of the 15 selected genes were considered suitable for further research.

2.2.5 qPCR

qPCR was performed for each primer pair and two candidate housekeeping genes with the 19 selected samples from the Gotland population. I used the same master mix composition and qPCR program as in section 2.2.4. Standards, cDNA template and negative controls were all loaded in three technical replicates.

2.2.6 Statistical analysis

Validation of housekeeping gene

Translation elongation factor alpha (eEF1α; FTEF) and α-tubulin were used as candidates for housekeeping genes (Rivera-Vega *et al.*, 2012; Sahraei, 2016). The primers are presented in Appendix 1. The genes were evaluated by inserting the standard curve efficiency and relative expression values into BestKeeper (Pfaffl), which gives information about standard deviation and Pearson correlation coefficient value.

Relative expression

Using the data from the qPCR, the efficiency for each gene and the mean of the three threshold cycle values for each individual were evaluated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) to calculate the relative expression.

Gene expression data comparison

In order to see whether the candidate genes showed the same pattern regarding higher expression in different phenotypes, I compared the differential expression of phenotypes between the present study and the previous study on *F. excelsior* (described in section 1.3.3).

One-way ANOVA on relative expression

One-way ANOVA (analysis on variation) was performed on each of the ten candidate genes, where relative expression between classification categories were tested for significant differences with Kruskal Wallis. A Mann Whitney U-test was thereafter carried out on genes showing a difference between either phenotype (MT and MS) or genotype (HT and Ht). Since the data was not normally distributed, I considered the Mann Whitney U-test to be appropriate.

3 Results

3.1 Validation of existing molecular marker

3.1.1 Sample collection

Of the 60 ash tree individuals initially classified based on phenotype in 2013, all retained their original classification in a second monitoring in autumn 2017. The classification was, therefore, considered reliable and all individuals were selected for this analysis.

3.1.2 Gel electrophoresis

The gel electrophoresis revealed primer dimers or degraded DNA in most reactions. As seen in Figure 3, some of the reactions did not give any product of the desired fragment and were, therefore, excluded from purification and sequencing.

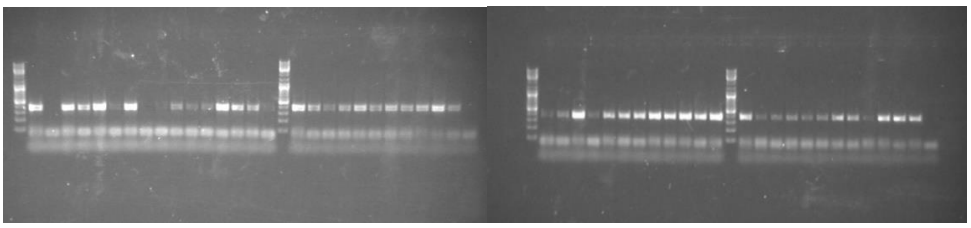


Figure 3. Gel image of cSNP amplicons. Ladder was loaded twice per gel.

3.1.3 Sequencing and analysis

Sanger sequencing results from Macrogen, analysed with SeqMan, are summarised in Table 1. 41 out of 48 samples were successfully sequenced, of which 34 were

classified as tolerant and seven were classified as susceptible according to Menkis' phenotyping (see section 2.1.1). 28 (82.35%) of the individuals with a tolerant phenotype carried the allele associated with reduced susceptibility, while four (57.14%) of the trees phenotypically classified as susceptible lacked the allele. 78.05% of the individuals had a corresponding phenotype and genotype (MTHT and MSHt). No samples were re-sequenced.

Table 1. *Number of sequenced individuals in each category for genotype (cSNP) and phenotype.*

		Genotype (Harper)			
		Tolerant	Susceptible	Unclear*	Total
Phenotype (Menkis)	Tolerant	28	6	6	40
	Susceptible	3	4	1	8
	Total	31	10	7	

*Sequencing result missing or not reliable.

Genotype frequencies for the phenotypically tolerant individuals were not consistent with Hardy-Weinberg equilibrium, while the individuals that were susceptible phenotypically were in Hardy-Weinberg equilibrium.

3.2 Relative expression of candidate defence genes

3.2.1 Sample selection

19 individuals were selected based on their genotype (absence or presence of the allele associated with reduced susceptibility presented by Harper *et al.* (2016)) and phenotype (see Table 2). Due to few samples with a successful sequencing with the phenotype classification susceptible (MS), I chose to include an individual with unclear genotype (MSHx).

Table 2. *Number of selected individuals in each category for genotype (cSNP) and phenotype.*

		Genotype (Harper)			
		Tolerant	Susceptible	Unclear*	Total
Phenotype (Menkis)	Tolerant	6	6	-	12
	Susceptible	3	3	1	7
	Total	9	9	1	

*Sequencing result not reliable.

3.2.2 Selection of candidate genes

15 genes were selected with the aim of successful analysis of approximately ten genes. The candidate genes are shown in Table 3, together with the relative expression data from the study described in section 1.3.3.

Table 3. *Candidate genes. Tolerant (To), susceptible (Su).*

Gene ID	Expression greater in (To or Su)	Relative expression To/Su	Annotation of best BlastX hit	Pathway
XLOC_002344	To	2.38044	lipoxygenase homology domain-containing protein 1-like	Hormone
XLOC_007773	To	1.49092	maternal effect embryo arrest 14 isoform 1	
XLOC_024086	Su	-1.88089	mads-box protein svp	TF
XLOC_028584	To	2.17356	21 kda	
XLOC_048502	Su	-2.10186	nac domain-containing protein 72	TF
XLOC_056239	To	1.37837	myb-related protein 308-like	TF
XLOC_057555	Su	-1.44922	myb-related protein myb4-like	TF
XLOC_059127	To	2.45966	kda class i heat shock	
XLOC_065948	To	2.36393	probable lrr receptor-like serine threonine-protein kinase at1g56140	LRR
XLOC_071076	To	1.80144	cbl-interacting serine threonine-protein kinase 6-like	LRR
XLOC_083555	To	1.35903	secoisolariciresinol dehydrogenase-like	Terpene
XLOC_083556	To	1.40586	secoisolariciresinol dehydrogenase-like	Terpene
XLOC_035278	To	1.47082	sucrose synthase 1	CAZY
XLOC_042290	To	1.53453	non-specific lipid-transfer protein 2-like	PR
XLOC_012264	To	2.19733	universal stress protein a-like protein	

3.2.3 Primers

All primers designed are shown in Appendix 1.

Primer testing - gel electrophoresis

Images of the gel are shown in Figure 4. Primers for gene XLOC_057555 did not amplify any product and primers for gene XLOC_012264 resulted in multiple bands. These genes were excluded from further analysis.

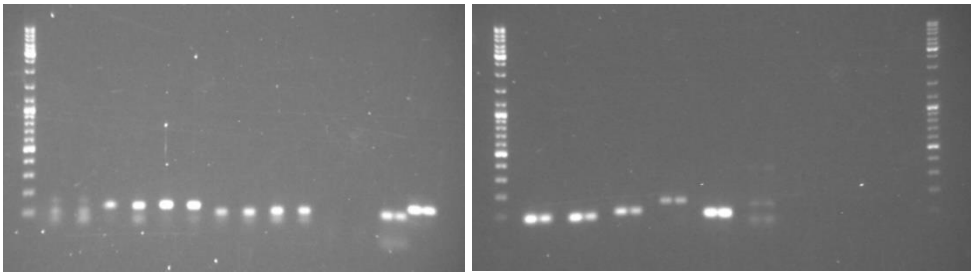


Figure 4. Gel images of primer amplicons. In the left image, the gap shows that gene XLOC_057555 resulted in no product. In the right image, the last wells with several products represent primers for gene XLOC_012264. Two negative controls were loaded next to XLOC_012264 and did not contain any product.

Primer testing - standards

After evaluating the qPCR results, ten of the 15 candidate genes were approved for use in the gene expression analysis, see Table 4.

Table 4. *qPCR results for standards.*

Gene ID	Efficiency (%)	R^2	Slope	Assay
XLOC_002344				No, qPCR failed.
XLOC_007773	104.4	0.966	-3.250	Yes
XLOC_024086	95.6-101.3	0.993	-3.433	Yes
XLOC_028584	86.7	0.993	-3.639	Yes
XLOC_048502				No, qPCR failed.
XLOC_056239	83.0	0.984	-3.811	Yes
XLOC_057555				No, no product in PCR.
XLOC_059127	107.0	0.978	-3.215	Yes
XLOC_065948	115.7	0.991	-2.951	Yes
XLOC_071076	96.5	0.995	-3.471	Yes
XLOC_083555	82.5-88.6	0.991	-3.826	Yes
XLOC_083556	93.9	0.995	-3.511	Yes
XLOC_035278				No, qPCR failed.
XLOC_042290	115.9	0.990	-2.978	Yes
XLOC_012264				No, several products in PCR.

3.2.4 qPCR

qPCR results for the ten candidate genes are presented in Appendix 2.

3.2.5 Statistical analysis

Housekeeping gene

According to BestKeeper (Pfaffl), FTEF was more reliable as housekeeping gene compared to α-tubulin. FTEF showed lower variance in expression across all samples, making it a more consistent baseline, and all data for relative expression was therefore calculated in relation to FTEF.

Relative expression

Values for relative expression were calculated using using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). The threshold cycle (C_T) values for each individual were mean values of the triplicate reactions. The data is presented in Appendix 2.

Gene expression data comparison

In the comparison between differential expression of phenotypes in the two studies (the present study and the previous study described in 1.3.3), the studies agreed for seven of ten genes, but for the remaining three genes the studies suggested differential expression in different directions (see Table 5).

Table 5. *Differential expression of phenotypes. Comparison between present study (leaf tissue) and previous study (bark tissue).*

	Expression greater in (To or Su)	
Gene ID	Bark	Leaf
XLOC_007773	To	To
XLOC_024086	Su	Su
XLOC_028584	To	To
XLOC_056239	To	Su
XLOC_059127	To	To
XLOC_065948	To	Su
XLOC_071076	To	To
XLOC_083555	To	To
XLOC_083556	To	To
XLOC_042290	To	Equal

One-way ANOVA

The one-way ANOVA exposed individual X007 (MSHx) as an outlier and it was therefore excluded from further analysis, leading to the removal of the classification MSHx.

None of the candidate genes had a p-value <0.1 in the Kruskal Wallis test, but six of them showed a trend (p-value <0.5), see Table 6. The test revealed a difference between the Menkis classifications (MT/MS) in five genes (XLOC_007773, XLOC_059127, XLOC_071076, XLOC_083555, XLOC_083556) and the Harper classifications (HT/Ht) in one gene (XLOC_042290), see graphs in appendix 3. In a Mann Whitney U-test, four of these were trending (p-value <0.1), one of which showing a significant p-value <0.05 as shown in Table 7 and graphs in Appendix 4.

Table 6. *P-values for Kruskal Wallis test for the ten candidate genes.*

Gene ID	Kruskal Wallis p-value
XLOC_007773	0.6546
XLOC_024086	0.3381
XLOC_028584	0.5895
XLOC_056239	0.2483
XLOC_059127	0.4949
XLOC_065948	0.514
XLOC_071076	0.1086
XLOC_083555	0.5242
XLOC_083556	0.4864
XLOC_042290	0.4095

Table 7. *P-values for Mann Whitney U-test for the six candidate genes that showed a difference in relative expression between phenotype or genotype classification categories. The asterisk indicates a significant p-value (<0.05) for gene XLOC_071076.*

	Mann Whitney U-test p-value	
Gene ID	MT/MS	HT/Ht
XLOC_007773	0.173	
XLOC_059127	0.0942	
XLOC_071076	0.0176*	
XLOC_083555	0.1067	
XLOC_083556	0.079	
XLOC_042290		0.0661

4 Discussion

4.1 Validation of cSNP

The cSNP Gene_22343_ Predicted_mRNA_scaffold3139:2378, identified as a moderately good predictor for reduced susceptibility to ash dieback by Harper *et al*. (2016), shows the same tendency in this study. 78% of the successfully sequenced individuals had a genotype that corresponded to the phenotype. Yet, to ensure this statement a larger sample size is required. Out of the 60 trees selected for the study, 48 cSNPs were sequenced. The sequencing results gave clear chromatograms for 41 individuals, while the rest either failed or had background noise in the chromatograms making them difficult to analyse. With more time available, these samples could have been re-sequenced.

The mismatches between phenotype and genotype can be due to errors in the PCR or qPCR assays, or other laboratory related conditions. More likely, they are connected to the fact that traits such as tolerance against pathogens are complex and prone to be additive, and the efficiency of the cSNP can be affected by additional genes, as stated by Harper *et al*. (2016). Since the cSNP is within a transcription factor gene, there might be genes downstream that induce tolerance. There is a chance that individuals have simply escaped infection, although I consider this unlikely since the selected trees in the Gotland population were located in stands where the surrounding trees were infected and *H. fraxineus* is wind-borne.

It was expected that the genotype frequencies for the phenotypically tolerant individuals would deviate from Hardy-Weinberg equilibrium, given that the trees were selected based on an expected heterozygosity for A and G alleles. The fact that the genotype frequencies for the susceptible phenotypes were consistent with Hardy-Weinberg equilibrium, indicates that the G allele is necessary but not sufficient for

the tolerance trait. Important to keep in mind, though, is that stochasticity increases with small sample sizes. The lack of homozygosity for the G allele also indicates that the A allele carries important traits. Nevertheless, the capacity of the cSNP identified by Harper *et al*. (2016) has limits.

4.2 Identification of molecular marker using qPCR

Out of 15 selected candidate genes, ten were analysed using qPCR. The remaining primers failed to either amplify the correct product in the PCR, or in the qPCR. The experiment could be repeated with optimised primers.

The one-way ANOVA showed an individual (X007, class MSHx) as an outlier and it was therefore excluded from further analysis. This can probably be explained by a low quality in the analysis with Agilent 2100 bioanalyzer (Agilent Technologies), RIN 4.7, and a weak fragment band in the gel electrophoresis as seen in Figure 5. The sample, however, was sent for sequencing, as I wanted as many phenotypically classified susceptible trees as possible for the analysis.

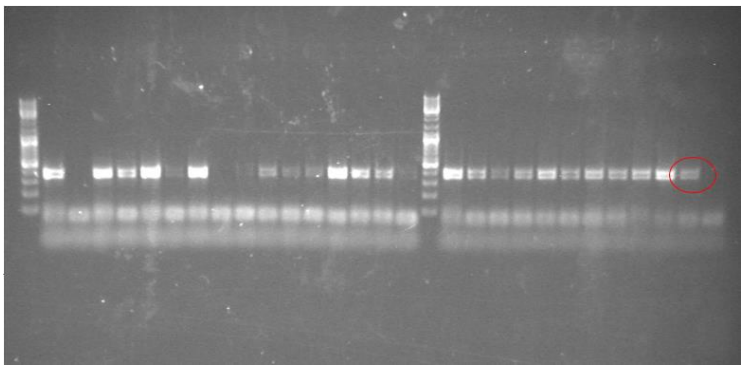


Figure 5. Gel image exposing a weak fragment band for individual X007 (indicated by red circle).

In the one-way ANOVA, none of the candidate genes had significant differential expression between tolerant and susceptible ash tree individuals. Again, a larger sample size would increase the reliability for these results. The Mann Whitney U-test showed four genes to be trending (p-value <0.1), and it is possible that these genes would show a higher significance if more material had been used in the study. However, one of the genes showed a significant p-value (<0.05), making it a strong candidate for further investigation. The gene carries a LRR-motif, an extracellular domain that is present in receptors within the innate immune system. The innate immune system constitutes the first line of defence against pathogens foreign to the

organism (Berg *et al.*, 2015). In fact, most resistance genes in crops carry LRR-motives (Malin Elfstrand, personal conversation).

It would be relevant to test if the same patterns appear in other LRR-motif carrying genes within the *F. excelsior* genome. Also, it would be interesting to further investigate the identified LRR-motif carrying gene in this study regarding its pathogen specificity.

If I were to do this study again I would devote more time evaluating the candidate genes and the samples selected. I missed out on one of the individuals with a corresponding phenotype and genotype for susceptibility, and since there were very few phenotypically susceptible trees in the experiment, adding more trees would have made the results for this category more reliable. Some of the qPCR efficiencies were on the border to being too high or too low. However, since all samples were expressed on a level such that the efficiency did not cause a major problem, I chose to include them anyway. Primers and standard dilutions were taken from a previous in-house study (Sahraei, 2016), and had therefore been stored in a freezer for several years, which might have affected the quality.

Relative expression is sensitive to several factors, such as the tree's age, the condition of the tree, etc. Though all trees in the Gotland population were mature, no factors are specified beyond their phenotype regarding dieback-related symptoms. The different directions regarding differential expression, showed in the comparison between this study and the previous study by Sahraei (2016), can be explained by tissue specific expression, since bark tissue was used in the first study while I extracted RNA from young leaves. Although, the data has low power in both studies and stochasticity probably affects the results. By taking several samples from each individual, from different tissues, tissue specific expression can be investigated.

The heat map in Appendix 2 reveals larger variances in relative expression within the phenotype classes than between them. As mentioned before, more material would possibly change that pattern. I would suggest that this experiment is repeated in a more controlled environment, with a larger sample size of clones from phenotypically classified individuals, preferably equally distributed between phenotype classifications.

4.3 Future perspectives

The genotypes associated with reduced susceptibility are not fully sufficient. There seem to be multiple factors that affect the trait, and a more thorough analysis of these factors is needed to find highly tolerant genotypes.

Combining the identified markers and further investigate genes related to them can provide us with more efficient and reliable genetic variants with high tolerance to ash dieback.

Apart from the epidemic of ash dieback, European forests are suffering from great losses of elm, alder and oak due to *Phytophthora* species (Pautasso *et al.*, 2013). Unfortunately, the *Fraxinus* species in Europe are expected to face major challenges in the near future, as the expansion rate of the wood-boring beetle emerald ash borer is fast and its impact devastating. The beetle is native to East Asia and has in the last decades caused great damage to ash species in North America. It was recorded in Moscow in 2003 and is expected to reach Central Europe within 15-20 years. All *Fraxinus* species in Europe and North America are presumed to be susceptible to emerald ash borer attacks, and the tree mortality is high (Valenta *et al.*, 2016).

The losses in European forests affect a massive number of organisms, and to inhibit the damage process it is important to identify trees that can withstand these threats and maintain a healthy, genetically diverse population. Molecular markers for prediction of tolerance to pathogens in trees can help develop sustainable strategies for conservation biology.

References

Agrios, G. N. (2005) *Plant Pathology*. 5th edn. London: Academic Press.

Artfakta (2015). Available at: http://artfakta.artdatabanken.se/taxon/220785 (Accessed: 3 May 2018).

Berg, J. M. *et al.* (2015) *Biochemistry*. 8th edn. Edited by L. Schultz. New York: W.H.Freeman Co Ltd.

Chen, C.-Y. (2014) 'DNA polymerases drive DNA sequencing-by-synthesis technologies : both past and present', 5(June), pp. 1–11. doi: 10.3389/fmicb.2014.00305.

Cleary, M. *et al.* (2016) 'Friend or foe? Biological and ecological traits of the European ash dieback pathogen Hymenoscyphus fraxineus in its native environment', *Scientific Reports*. Nature Publishing Group, 6(February), pp. 1–11. doi: 10.1038/srep21895.

Cleary, M. R. *et al.* (2014) 'Genotypes of Fraxinus excelsior with different susceptibility to the ash dieback pathogen Hymenoscyphus pseudoalbidus and their response to the phytotoxin viridiol - A metabolomic and microscopic study', *Phytochemistry*. Elsevier Ltd, 102, pp. 115–125. doi: 10.1016/j.phytochem.2014.03.005.

Griffiths, A. J. *et al.* (2000) 'An Introduction to Genetic Analysis', in *Mapping with molecular markers*. 7th edn. New York: W. H. Freeman.

Gross, A. *et al.* (2014) 'Corrigendum to "Reproductive mode and life cycle of the ash dieback pathogen Hymenoscyphus pseudoalbidus" [Fungal Genet. Biol. 49 (2012) 977-986]', *Fungal Genetics and Biology*. Elsevier Inc., 67, p. 82. doi: 10.1016/j.fgb.2014.03.004.

Harper, A. L. *et al.* (2012) 'Associative transcriptomics of traits in the polyploid crop species Brassica napus', *Nature Biotechnology*. Nature Publishing Group, 30(8), pp. 798–802. doi: 10.1038/nbt.2302.

Harper, A. L. *et al.* (2016) 'Molecular markers for tolerance of European ash (Fraxinus excelsior) to dieback disease identified using Associative Transcriptomics', *Scientific Reports*, 6(January), pp. 1–7. doi: 10.1038/srep19335.

Johanson, S. (2009) 'Svampen bakom askskottsjukan', *Fakta Skog*, (3).

Klug, W. S. *et al.* (2014) *Essentials of Genetics*. 8th edn. London: Pearson Education Ltd.

Kover, P. X. and Schaal, B. A. (2002) 'Genetic variation for disease resistance and tolerance among Arabidopsis thaliana accessions', *Proceedings of the National Academy of Sciences of the United States of America*, 199(17), pp. 11270–11274. doi: 10.1073/pnas.102288999.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method', *Methods*, 25(4), pp. 402–408. doi: 10.1006/meth.2001.1262.

Lopez-Pardo, R. *et al.* (2013) 'Validation of molecular markers for pathogen resistance in potato', *Plant Breeding*, 132(3), pp. 246–251. doi: 10.1111/pbr.12062.

Marigo, G. *et al.* (2000) 'Success in the demographic expansion of Fraxinus excelsior L.', *Trees - Structure and Function*, 15(1), pp. 1–13. doi: 10.1007/s004680000061.

More, D. (2013) *Illustrated Trees of Britain and Northern Europe*.
Pautasso, M. *et al.* (2013) 'European ash (Fraxinus excelsior) dieback - A conservation biology challenge', *Biological Conservation*. Elsevier Ltd, 158, pp. 37–49. doi: 10.1016/j.biocon.2012.08.026.
Rivera-Vega, L. *et al.* (2012) 'Evaluation of Reference Genes for Expression Studies in Ash (Fraxinus spp.)', *Plant Molecular Biology Reporter*, 30(1), pp. 242–245. doi: 10.1007/s11105-011-0340-3.
Sahraei, S. E. (2016) 'Gene expression patterns in Fraxinus excelsior clones with contrasting susceptibility to Hymenoscyphus fraxineus'.
Scherrer, D., Bader, M. K.-F. and Körner, C. (2011) 'Drought-sensitivity ranking of deciduous tree species based on thermal imaging of forest canopies', *Agricultural and Forest Meteorology*, 151(12), pp. 1632–1640. doi: 10.1016/j.agrformet.2011.06.019.
Stener, L. G. (2018) 'Genetic evaluation of damage caused by ash dieback with emphasis on selection stability over time', *Forest Ecology and Management*. Elsevier, 409(December 2017), pp. 584–592. doi: 10.1016/j.foreco.2017.11.049.
Stenlid, J. *et al.* (2017) 'Genomes of Hymenoscyphus fraxineus and Hymenoscyphus albidus encode surprisingly large cell wall degrading potential, balancing saprotrophic and necrotrophic signatures', *Baltic Forestry*, 23(1), pp. 41–51.
Thermo Fisher Scientific DNA Copy Number and Dilution Calculator (no date). Available at: https://www.thermofisher.com/se/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/dna-copy-number-calculator.html#/legacy=www.thermoscientificbio.com (Accessed: 17 April 2018).
Untergasser, A. *et al.* (2018) 'Primer3 — new capabilities and interfaces', 40(15), pp. 1–12. doi: 10.1093/nar/gks596.
Valenta, V. *et al.* (2016) 'A new forest pest in Europe : a review of Emerald ash borer (Agrilus planipennis) invasion', pp. 507–526. doi: 10.1111/jen.12369.

Acknowledgements

I would like to thank my supervisor Malin for all the support, patience and knowledge she has given me. I would also like to thank my assistant supervisors; Audrius, who has provided me with plant material and phenotype data, and Eddy, who has answered more questions than Google during this project. I am very thankful for all the support the people in the department of Forest Mycology and Plant Pathology have given me. Special thanks to my friends Helen, Eryn and Liam, for knowing their native language so well, and to Anna, Laura, Krystal, Marie and Lena for always pushing me through my stalls and coping with me in good times and bad.

Appendix

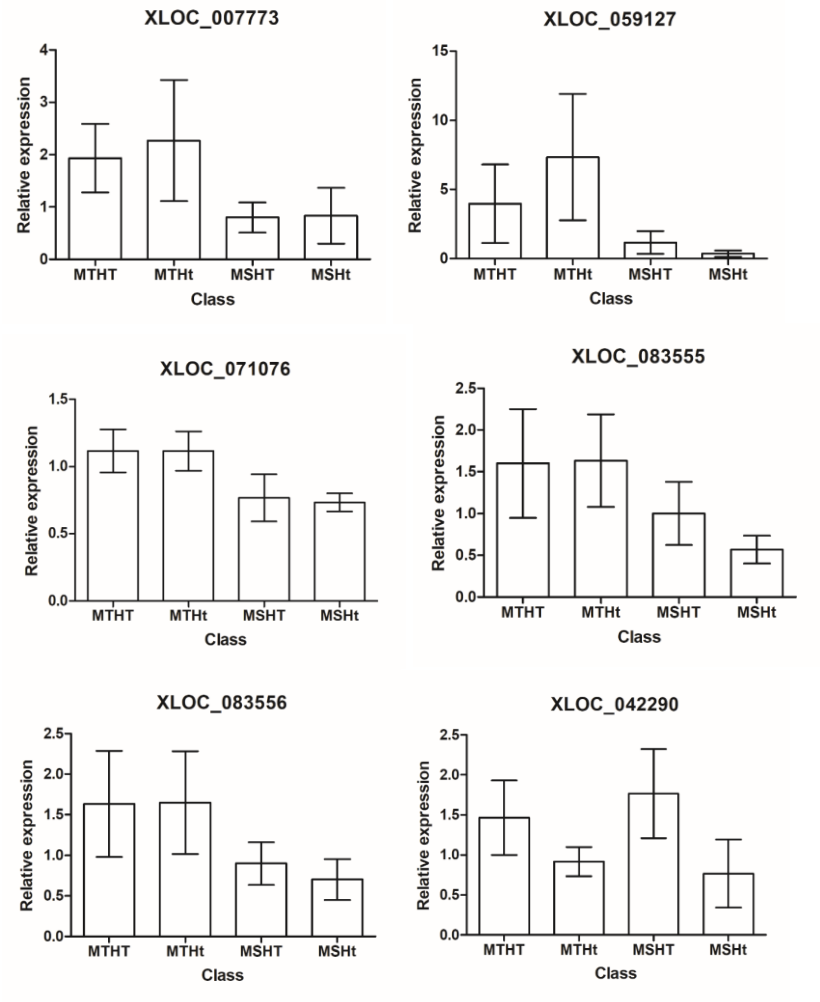
Appendix 1. Primers used for sequencing and qPCR.

Primer	Sequence	Annealing temperature (°C)	Product size (bp)	Type
Gene_22343-F	GGTTTCTCTTCTGCAGCGAG	59.4		
Gene_22343-R	TCCATGATCATCTTGCTGAG	55.3	230	cSNP
α-tubulin F	CACCTCCTCCAACGGTCTTA	53.8		
α-tubulin R	GGCTGGTATTCAGGTTGGAA	51.8	104	Housekeeping gene
FTEF-F	ACCAGCAAGTCCCAGTTGAGATG	62.4		
FTEF-R	TGAGCCAGGTTCAGCTTCCAATG	62.4	77	Housekeeping gene
XLOC_002344-F	CATCGTCACCTCCACGTAGTT	60.04		
XLOC_002344-R	AGAAGCCCATGTTTGTCCAG	60.11	102	qPCR
XLOC_007773-F	ATACCGAAGCTGGAGCCATT	60.98		
XLOC_007773-R	CTGCGAGTTGGGTTTCTGAT	60.25	91	qPCR
XLOC_024086-F	ATCGTCCCTTCGTCTGTTTG	60.11		
XLOC_024086-R	CTTCTCCCTCGCCATGTTTA	60.21	115	qPCR
XLOC_028584-F	GCCCGACTCACAAAATTCAG	60.64		
XLOC_028584-R	GCTCTTCAGCTCCTTGATCG	60.24	120	qPCR
XLOC_048502-F	TGTGCCCAACAAACAAAATG	60.39		
XLOC_048502-R	CTTGCGATGGGAGTGTGATT	61.07	88	qPCR
XLOC_056239-F	GCCGGAAGATTACCAGGAAG	60.95		
XLOC_056239-R	GGATCAATACCTCGGCTCAA	60.04	92	qPCR
XLOC_057555-F	TGAGACCTTTTGGTCCAGTGA	60.67		
XLOC_057555-R	AGACGAAGGGAATCCGAACT	60.07	106	qPCR
XLOC_059127-F	TGAGGAAGGGAACATTCTGC	60.2		
XLOC_059127-R	CGGTGCCAGTTATCGTTCTT	60.13	75	qPCR
XLOC_065948-F	GTCACTTCAAAGCCCGGATA	60.07		
XLOC_065948-R	GTTTGATGGATTGCCCTGAC	60.33	117	qPCR
XLOC_083555-F	ACATGGGGTGGTAGGACTCA	60.24		
XLOC_083555-R	GTGGTGACACGCAGTTTACG	60.22	80	qPCR
XLOC_083556-F	TGAGTCCTACCACCCCATGT	60.24		
XLOC_083556-R	TCGCAGTGGGAACATAATCA	60.07	101	qPCR
XLOC_035278-F	ATCGGCCATTAGTTCAGCAG	60.24		
XLOC_035278-R	TTTGGACTTACGGTGGTTGAG	60.02	150	qPCR
XLOC_042290-F	CACTCAGCCCTTGCATGTC	60.42		
XLOC_042290-R	AAGGCTGCTGCTCCTTGAT	60.11	84	qPCR
XLOC_012264-F	CCTCCTCTGGATGTCACTGG	60.67		
XLOC_012264-R	GCGACAGTTGCGTTGAAGT	60.05	134	qPCR

Appendix 2. qPCR results. Efficiency and heat map of relative expression data. Colours ranging from blue (low relative expression) to red (high relative expression).

Class	ID	Gene									
		XLOC_059127	XLOC_065948	XLOC_071076	XLOC_083555	XLOC_083556	XLOC_007773	XLOC_024086	XLOC_028584	XLOC_042290	XLOC_056239
Efficiency (%)		107.0	115.7	96.5	82.5-88.6	93.9	104.4	95.6-101.3	86.7	115.9	83.0
MTHT	F029	0,4	0,5	0,9	0,4	0,4	2,3	1,3	0,5	1,2	0,6
	F033	1,0	5,2	0,9	1,8	2,2	3,6	0,8	0,5	0,3	0,7
	F039	3,0	0,4	1,0	2,0	1,4	0,4	0,6	1,1	0,3	1,0
	F086	1,3	3,2	1,1	0,4	0,6	0,6	0,6	0,8	2,0	1,1
	F111	18,0	1,4	1,9	4,5	4,6	4,0	1,4	2,2	3,3	2,3
	F124	0,1	0,2	0,9	0,5	0,6	0,7	1,7	1,3	1,7	0,5
MTHt	F135	0,6	2,9	0,9	0,5	0,5	0,6	0,6	1,4	1,4	0,3
	F097	15,6	0,4	1,5	3,0	1,9	6,4	1,1	3,4	0,6	0,5
	F102	26,5	0,4	0,7	3,4	4,6	5,4	1,2	1,1	0,3	0,6
	F085	0,7	0,5	1,1	0,4	0,6	0,2	1,4	0,8	1,3	7,9
	F057	0,2	0,4	0,9	0,5	0,7	0,3	1,0	0,3	0,7	0,6
	F034	0,4	0,7	1,6	2,0	1,6	0,7	0,8	1,5	1,2	0,6
MSHT	X04	2,8	0,7	1,1	1,7	1,3	0,8	0,6	1,1	2,8	3,0
	X01	0,5	5,2	0,5	0,4	0,4	0,3	0,8	0,7	1,6	0,8
	X06	0,2	0,9	0,7	0,9	1,0	1,3	1,3	0,4	0,9	3,7
MSHt	X02	0,8	0,4	0,8	0,4	0,5	0,3	1,1	0,6	1,6	2,0
	X03	0,1	5,3	0,8	0,4	0,4	0,3	1,5	1,1	0,2	0,8
	X10	0,2	0,5	0,6	0,9	1,2	1,9	1,7	0,9	0,5	0,5
MSHx	X007	6,5	3,5	3,4	2,9	1,3	2,9	0,8	5,2	5,7	1,1

Appendix 3. Graphs showing the relative expression for the six genes that indicated a difference between either the phenotype classifications (XLOC_007773, XLOC_059127, XLOC_071076, XLOC_083555, XLOC_083556), or the genotype classifications (XLOC_042290).



Appendix 4. Graphs showing the results of Mann Whitney U-test for the four genes with a p-value <0.1. The asterisk indicates a significant p-value (<0.05) for gene XLOC_071076.