Genomics Approach to Understanding Silicon in Ash Trees (*Fraxinus excelsior*)

James Andrew Bedford

Master of Science (by Research)

University of York

Biology

September 2020

Abstract

Silicon is a non-essential plant nutrient thought to be beneficial for improving resistance against a wide range of herbivores and pathogens. However, the genetics underlying silicon uptake, deposition and disease resistance are not fully understood, especially in tree species such as Common ash (*Fraxinus excelsior*), which is severely affected by the ash dieback epidemic.

Therefore, the project aims to improve our knowledge of silicon in ash trees using measurements of silicon in branch and leaf tissue. It also aims to identify gene markers for silicon content and understand the genetics associated with silicon. Finally, it aims to improve the genetic resources available for ash through genome assembly with long-read data.

Whole-genome sequencing and assembly resulted in a highly contiguous genome, using fewer contigs than the previous 2017 assembly. Realignment of cDNA sequences was used to update gene order, which was implemented into the associative transcriptomic (AT) analysis to visualise the results.

AT analysis of branch silicon data and differential gene expression analysis of leaf silicon revealed 9 genes significantly associated with branch silicon content and 22 genes with differential expression between trees with high or low leaf silicon levels. A diverse range of functions are predicted for these genes, with several functioning in the growth and development of leaf and root tissue. The differential gene expression analysis also identified three defence-associated genes and two secondary metabolite biosynthesis genes with possible pest repellent and antifungal activity, based on the functions of these metabolites in other species.

The results from the genome-wide silicon analyses provide a framework for additional studies investigating the genetics of silicon in ash and its association with disease resistance. Additionally, the production of an improved *F. excelsior* genome will assist future genetic studies of ash trees.

List of contents

Abstract 2

List of contents 3

List of tables 5

List of figures 6

List of accompanying material 8

Acknowledgements 8

Declaration 8

Chapter 1 - Introduction 9

1.1 Threats to ash 9

1.2 Silicon in defence 11

1.3 The silicon cycle 14

1.4 Silicon accumulation in plants: uptake and deposition 16

1.4.1 Silicon uptake 16

1.4.1 Silicon deposition 22

1.5 Project aims 26

Chapter 2 – Production of an improved *F. excelsior* genome 27

2.1 Introduction 27

2.2 Methods 29

2.2.1 DNA extraction, sequencing and assembly 29

2.2.2 Assembly analysis and quality control 30

2.2.3 Identification of mitochondria and chloroplast contigs 32

2.2.4 Implementing transcriptome data into the ash genome 32

2.3 Results 33

2.3.1 Production of a more contiguous *F. excelsior* genome using long-read Nanopore sequencing 33

2.3.2 Identification of mitochondria and chloroplast genomes 36

2.3.3 Remapping cDNA sequences to improve gene position for transcriptome-based genetic analyses 37

2.4 Discussion 39

Chapter 3 – Physiology of silicon in ash 43

3.1 Introduction 43

3.2 Methods 45
3.2.1 Ash leaf sampling and silicon analysis 45
3.3 Results 47
3.3.1 Ash tree leaves accumulate silicon over time 47
3.3.2 Branch tissue accumulates significantly less Si than leaves 48
3.4 Discussion 51
Chapter 4 – Associative transcriptomics analysis of branch silicon data 54
4.1 Introduction 54
4.2 Methods 55
4.2.1 Field site and branch sample collection 55
4.2.2 Branch sample processing and XRF silicon analysis 56
4.2.3 Associative transcriptomics of branch Si traits 56
4.2.4 Bioinformatic analysis of genetic markers 58
4.3 Results 59
4.3.1 Associative transcriptomics reveals genetic markers associated with branch Si content 59
4.4 Discussion 66
Chapter 5 – Differential gene expression analysis of leaf Si data 72
5.1 Introduction 72
5.2 Methods 74
5.2.1 Sample selection and DNA extraction 74
5.2.2 Differential gene expression analysis 75
5.3 Results 76
5.3.1 Identification of differentially expressed genes between high and low leaf Si groups 76
5.4 Discussion 82
Chapter 6 – Project discussion and conclusions 88
Abbreviations 93
References 94

List of tables

Table 2.1. SeqKit statistics of the genome assembly pre- and post-filtered. MRD = median read depth. (Page 36)

Table 2.2. Comparison of the existing published genome from Sollars *et al*. (2017) and the genome produced from this project. (Page 36)

Table 4.1. Ash dieback disease severity score key. (Page 55)

Table 4.2. GEMs associated with the Si relative to branch diameter (% Si mm^{-1}) trait, passing FDR and Bonferroni thresholds. (Page 59)

Table 4.3. GEM associated with the branch Si content (%) trait, passing FDR and Bonferroni thresholds. (Page 60)

Table 5.1. Differentially expressed genes between high and low silicon groups (Log2 fold change >1 and adjusted p value < 0.05). Annotations from BLASTx of the gene sequence. SE is the standard error of the log2-fold change. (Page 77)

Table 5.2. Differentially expressed genes between high and low silicon groups (Log2 fold change < -1 and adjusted p value < 0.05). Annotations from BLASTx of the gene sequence. SE is the standard error of the log2-fold change. (Page 81)

List of figures

Figure 1.1. Model of silicic acid [$Si(OH)_4$] uptake from the soil solution and transport along the roots to the stele/xylem, illustrating the differences in root structure and silicon transporter (LSI) localisation between maize/barley and rice. Figure from Mitani *et al.* (2009). (Page 21)

Figure 2.1. Unfiltered genome assembly. Each point is a contig, separated by length and percentage GC content. The red line is the expected nuclear genome percentage GC content (34%) and the blue line is the expected mitochondrial GC content (45%). (Page 31)

Figure 2.2. Method used to filter the genome using uniqueness and median read depth (MRD) properties. Each point represents a contig, coloured by contig length on a log10 scale. The shaded region contains the contigs filtered out of the genome after applying the filter: MRD > 6 OR (Unique >= 50% AND 1 <= MRD <= 6). (Page 34)

Figure 2.3. Number of alignments to the genome per cDNA, according to GMAP. A total of 50841 cDNAs were analysed. (Page 38)

Figure 3.1. Analysis of leaf silicon data. **(a)** Leaf silicon content of samples collected in August from trees either fertilised with potassium silicate or left unfertilised (n=10 per group). **(b)** percentage change in leaf silicon between samples collected in July and August. **(c)** leaf Si measurements from July and August samples. (Page 47)

Figure 3.2. Analysis of branch and leaf Si data. **(a)** Variation in branch Si content across 125 ash trees. **(b)** Comparison between branch and leaf Si levels. **(c)** Leaf to branch Si ratio (n = 103). Each accession, n, is a mean of three biological replicates. (Page 49)

Figure 3.3. Linear regression between branch diameter and branch Si content. n = 374. Plot annotated with regression line equation and R^2 value. (Page 50)

Figure 3.4. Branch Si content found in ash trees with a range of ash dieback disease severity scores. Boxplots annotated with a different letter are significantly different. Sample counts are 88, 70, 68, 72 and 121, for disease scores 1 to 5, respectively. (Page 50)

Figure 4.1. Field site images of trees with an ash dieback disease severity score of 1 and 5. (Page 55)

Figure 4.2. Transcript expression associated with the Si relative to branch diameter trait. The significance of trait association is displayed as a -$\log_{10}$(p) scale. Each point represents a gene and are pseudo-positioned based on their location in the genome, predicted in Chapter 2. Genome contigs are ordered alphabetically and are identified by alternating colours of grey/black for genes that do not pass FDR and green/orange for GEMs that pass FDR. Within contigs, genes are positioned based on their sequence mapping. Genes with multiple alignments are represented as symbol (**x**), and single mappings with (●). In the case of multiple alignments, only a single position is plotted. Genes with no significant alignments to the genome are located in the "NA" panel. The dashed line represents the Bonferroni multiple test correction threshold. The GEMs discussed in this chapter are labelled with their BLASTX annotation. (Page 63)

Figure 4.3. Transcript expression associated with the branch Si trait. The significance of trait association is displayed as a -$\log_{10}$(p) scale. Each point represents a gene and are pseudo-positioned based on their location in the genome, predicted in Chapter 2. Genome contigs are ordered alphabetically and are identified by alternating colours of grey/black for genes that do not pass FDR and green/orange for GEMs that pass FDR. Within contigs, genes are positioned based on their sequence mapping. Genes with multiple alignments are represented as symbol (**x**), and single mappings with (●). In the case of multiple alignments, only a single position is plotted. Genes with no significant alignments to the genome are located in the "NA" panel. The dashed line represents the Bonferroni multiple test correction threshold. The GEMs discussed in this chapter are labelled with their BLASTX annotation. (Page 64)

Figure 4.4. Transcript abundance (RPKM) of five GEMs (a-e) associated with branch silicon content. The green lines represent regression lines, calculated using linear models, and are labelled with their corresponding line equation and R^2 value. n = 114. (Page 65)

Figure 5.1. High and low leaf Si content grouping for differential gene expression analysis. Group means represented as black points and bars are +/- standard error. Grey points are individual values (n=5 per group). (Page 74)

List of accompanying material

Contig information from Tapestry

The Hawkhills ash leaf silicon data

Cemetery fields branch silicon data

Associative transcriptomics trait data

Acknowledgements

I would like to thank the Department of Biology Genomics and Bioinformatics Technology Facility at the University of York for providing support throughout the project. Specifically, Dr Sally James for assistance with genomic DNA extraction and for the Nanopore sequencing. As well as Dr John Davey and Dr Katherine Newling for guidance on genome assembly.

The *F. excelsior* genome assembly and analysis were undertaken on the Viking Cluster, which is a high-performance compute facility provided by the University of York. I am grateful for computational support from the University of York High Performance Computing service, Viking and the Research Computing team.

Finally, I thank my supervisors Dr Andrea Harper and Professor Sue Hartley for their guidance and support throughout the project.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1 - Introduction

1.1 Threats to ash

European or Common ash trees, *Fraxinus excelsior* (*F. excelsior*), are a valuable component of woodlands and plantations across the United Kingdom (UK) and the European mainland. The native range of this deciduous tree species spans across the European continent. In more northern regions, including the UK, trees are present in lowland mixed woodlands, shared with species such as European beech, oak and sycamore. In warmer climates, ash is found at higher altitudes in mountainous regions (reviewed by Dobrowolska *et al*., 2011).

F. excelsior is a dicotyledon and member of Oleaceae family, including species such as olive, privet and jasmine. It is a relatively fast-growing tree species, reaching maximum height 30 to 40 years earlier than beech trees (reviewed by Dobrowolska *et al*., 2011) and provides critical ecological services and economic resources. The hardwood timber of ash trees has desirable properties; it has a greater resistance to fractures compared to softwood conifer species, high longitudinal strength and elasticity (Niemz *et al*., 2014). The fast-growing nature and useful wood properties make it a valuable economic resource.

Ash trees are prevalent within mixed woodlands in the UK and are an integral component of these ecosystems. There are approximately 126 million ash trees across the UK, accounting for 14% of all broadleaf tree species in the UK (Forestry commission, 2013). A variety of insects, fungi and lichens are thought to rely on ash trees in the environment, including several moth species, butterflies and aphids (Biological Records Centre, 2008).

However, European and UK ash trees are currently suffering from the ash dieback epidemic caused by an invasion of the ascomycete fungus *Hymenoscyphus fraxineus* (*H. fraxineus*). The native range of the fungus spans across eastern Asia, with populations thought to originate from Japan (Zhao *et al*., 2012; McMullan *et al*., 2018). In this ecosystem, the fungus infects *Fraxinus mandshurica* (*F. mandshurica*), a native ash species, where it interacts biotrophically with the host. This interaction is asymptomatic and is non-pathogenic during the asexual phase of the fungus. Once host senescence and defoliation occur, *H. fraxineus* switches to its sexual phase and fruiting bodies begin to form (Inoue *et al*., 2019).

Due to this biotrophic interaction with the host ash in its native range, *H. fraxineus* is regarded as a decomposer and aids the cycling of nutrients from the leaves.

The invasive subpopulation of *H. fraxineus* has low genetic diversity due to a prominent founder's effect (McMullan *et al*., 2018), however, it has been highly successful in colonising and infecting European ash species, including European ash (*F. excelsior*) and narrow-leaf ash (*F. angustifolia*). It was first detected in the UK in 2012, although was likely present before this time. Since then, infected trees have been identified across the whole of the UK.

The infection begins when airborne ascospores spores contact an ash leaf. These spores germinate and appressorium develop, providing the force required to penetrate the leaf cuticle and cell walls (Cleary *et al*., 2013). Hyphae growth occurs within a single cell while the infection begins to establish. Once the hyphae begin to spread the fungus switches to necrotic phase and leaf lesions develop. The hyphae mycelium grows down the petiole and infects the leaf stem, eventually reaching the wood tissue in severe infections (Gross *et al.,* 2014). This results in progressive crown dieback, increased susceptibility to secondary infections and can lead to lesion formation in the bark at nodes connecting to the infected leaf stems (Gross *et al.,* 2014). In early summer, *H. fraxineus* switches to its sexual morph and fruiting bodies (apothecia) are matured between July and August (Cleary *et al*., 2013), followed by ascospore production and further transmission of the fungus.

Compared to the interaction with *F. mandshurica*, *H. fraxineus* appears to switch to a necrotrophic phase within approximately 4 days of infecting a leaf of *F. excelsior* (Mansfield *et al*., 2018). This change in virulence of the pathogen has caused significant damage to the UK ash population and to the UK economy and will continue to do so in the future. Models have predicted approximately 70% mortality rates for ash growing in woodlands and 85% for plantations (Coker *et al*., 2018). The severity of the epidemic will cost the UK £14.8 billion over the next 100 years, of which £7.8 billion will occur in the next 10 years, as a result of felling costs, replanting and damages to ecosystem services (Hill *et al*., 2019).

In addition to the damage caused by ADB, European ash species are likely to face an additional threat from Emerald ash borer (EAB) in the future.

EAB is an invasive beetle with wood-boring larvae and originates from China and eastern Russia (Herms and McCullough, 2014). EAB is dispersing towards Europe from Russia and is predicted to reach this region by 2030-2035 (Valenta *et al*., 2017).

EAB has already devastated North American ash species, including *Fraxinus pennsylvanica, F. americana* and *F. nigra*, after suspected colonisation through contaminated wooden pallets. Infected trees will display loss of canopy and branch dieback, followed by girdling as the wood-boring larvae disrupt the vascular tissue, resulting in the death of the tree. EAB is estimated to have costed the North American economy $10.8 billion over the last 10 years, including the costs of treating, removing, and replacing ash trees (Kovacs *et al*., 2010).

F. excelsior is a genetically diverse species and resistance to *H. fraxineus* and EAB is likely present in the population. However, ADB causes high mortality in *F. excelsior* populations and therefore creates a large bottleneck in diversity for tolerant populations. By restricting the diversity of the natural population, susceptibility to future invasions of *H. fraxineus* and EAB will increase, decreasing the probability of *F. excelsior* recovering from these epidemics.

1.2 Silicon in defence

Plants respond to biotic stresses, including fungal pathogens and insect herbivores, through several layers of defence. These involve physical, biochemical and inducible defences, that function to prevent and minimise pathogen infection. Physical defences have the primary interaction with pathogens and act as a protective barrier against infection. Lignin, silicon, the cuticle and cell wall are examples of these structural barriers and prevent cell penetration by pathogens. Lignin, for example, is most prevalent in the plant secondary cell wall, where it integrates into the cellulose and hemicellulose polymer network and increases the rigidity and hydrophobic properties of the cell wall. This is thought to improve resistance to fungal appressoria penetration and prevent the spread of pathogen-secreted degradative enzymes. Reduction in lignin production, through knockdown of lignin biosynthesis enzymes, is shown to increase susceptibility to fungal pathogens such as powdery mildew (Bhuiyan *et al*., 2009).

Biochemical defences tend to be non-specific and target groups of pathogens, rather than certain species. These can include phytoalexins, glucosinolates and defensins, which have been demonstrated to be utilised by plants against fungi, bacteria and herbivores (reviewed by Zhang *et al*., 2018). Conversely, inducible defences provide specific response against pathogen infection. Pathogen/microbe-associated molecular patterns (PAMPs/MAMPs), such as fungal chitin and flagellin, are detected by plant-expressed pattern recognition receptors (PRRs) including leucine-rich repeat (LRR) receptor kinases, potentially initiating pattern-triggered immunity (PTI). Detection also occurs through recognition of pathogen effectors, such as proteases, glycolipids and toxins, by disease resistance (R) proteins and can result in effector-triggered immunity (ETI) (reviewed by Bigeard *et al*., 2015). ETI and PTI can lead to the hypersensitive response, which is rapid and localised programmed cell death that can prevent the spread of the pathogen, and systemic acquired resistance (SAR), a whole-plant priming of pathogenesis-related (PR) genes (reviewed by Conrath, 2006).

Biological silica is hypothesised to provide general resistance to a range of pathogens, herbivores and insect pests. This has been demonstrated through numerous studies that highlight the beneficial effect of silicon fertilisation in increasing disease resistance in a variety of plant species. For example, silicon application can reduce the number of leaf blast lesions formed in rice when infected with the necrotrophic fungus *Magnaporthe grisea*. The increase in resistance against this pathogen was attributed to the accumulation of silicon in rice leaves, as the positive increase in resistance is reversed in rice mutants that were defective in silicon uptake and contained significantly lower silicon content (Nakata *et al*., 2008). Additionally, in banana plants, silicon supplementation has been demonstrated to increase silicon deposition and resistance to air-borne pathogens, such as the fungi *Mycosphaerella fijiensis* that causes black sigatoka (Kablan *et al*., 2012), and soil-borne pathogens including *Fusarium oxysporum* (Fortunato *et al*., 2014).

In addition to monocots, the beneficial effects of silicon in disease resistance has been illustrated in dicot species, including in olive trees, which are closely related to *F. excelsior* and share the same family. Silicon fertilisation using an orthosilicic acid solution resulted in a reduction in disease severity and fewer leaf lesions, following inoculation with the

deuteromycete fungal pathogen *Spilocaea oleaginea*, which causes olive leaf spot disease. The extent of leaf spot reduction was proportional to the concentration of silicon supplied to the trees in one of two cultivars tested (Nascimento-Silva *et al*., 2019).

The mechanisms of how the beneficial effect of silicon fertilisation are unclear. However, it is suspected that silicon may provide a physical barrier against pathogens, induce defence chemicals, modulate SAR and regulate transcriptional profiles (reviewed by Wang *et al.,* 2017; Jiang *et al*., 2019).

In bacterial wilt-susceptible tomato plants (*Solanum lycopersicum*), treatment with 2 mM potassium silicate increased resistance to the soil-borne bacterium *Ralstonia solanacearum* (*R. solanacearum*), improving the disease index by 64.5% after 7 days (Jiang *et al*., 2019). The silicon treatment appeared to upregulate ethylene and salicylic acid production after an initial delay, and increased jasmonic acid production for the first 3 days of infection. As ethylene is linked with response to necrotrophic pathogens, salicylic acid to biotrophic pathogens and jasmonic acid to herbivores (Bari and Jones, 2009), this could explain the link between plant silicon and disease resistance. It has also been demonstrated that when jasmonic acid signals are blocked, through knockdown of jasmonic acid biosynthesis and responsive genes, there is lower expression of the three major silicon uptake transporters OsLSI1, 2 and 6 and reduced leaf silicon content (Ye *et al*., 2013). Therefore, indicating a feedback system between silicon and jasmonic acid biosynthesis.

Silicon also appears to have an indirect role in the modulation of gene expression. Silicon treated tomato plants infected with *R. solanacearum* showed thousands of differentially expressed genes compared to untreated controls (Jiang *et al*., 2019). Differentially expressed genes encoded proteins with functions related to reactive oxygen species (ROS) and the phenylpropanoid biosynthesis pathway. As well as calcium and MAPK signalling pathway genes, PRRs, PR genes, SAR-related genes and stress-related genes. This illustrates the potential effects silicon has on the expression profiles of several disease response genes.

In addition to bacterial and fungal pathogens, silicon improves resistance to herbivorous insects and animals. Rice field trials have revealed that silicon fertilisation reduces plant damage by the rice stem borer *Chilo suppressalis* and the leafroller *Cnaphalocrocis medinalis* (Han *et al*., 2018). Silicon

treatment influenced the growth of stem borer larvae and decreased the weight of the pupae and larvae, indicating that silicon reduces the feeding of stem borer insects. Similar effects of silicon were observed for the insect herbivore *Spodoptera exempta*, which displayed increased mandible wear and reduced growth rate when restricted to a high silica diet (Massey and Hartley, 2008).

Based on these experiments, it appears that silicon can provide general resistance against a wide range of pathogens and insect pests, achieved through modulation of transcriptional profiles, signalling and interactions with the plant defence hormone systems. This makes silicon a potential solution for combined resistance to fungal pathogens, such as *H. fraxineus,* and insects like EAB.

1.3 The silicon cycle

Silicon is the second most abundant element in the Earth's crust, containing an average 28% Si by weight. The presence of Si is determined by rock type, with high concentrations found in sedimentary and igneous rocks consisting of amorphous Si or crystalline Si structures such as quartz. Whereas calcareous rocks including limestone contain much lower levels of silicon.

Comparable to other plant elemental nutrients, such as nitrogen and potassium, silicon is cycled through the environment. The major reservoirs of silicon in the cycle are plant material, the soil solution, minerals and other organic matter (Bartoli, 1983), with processes including plant-soil interactions, decomposition of plant tissue and weathering mechanisms connecting the Si reservoirs.

Silicon located in rocks and minerals represents the largest source of Si and is the initial supply of Si in the cycle. Chemical weathering is responsible for dissolving silicate minerals into a soluble form which enters the soil solution as monomeric orthosilicic acid ($Si(OH)_4$). This is the plant-accessible form of Si that it is deposited as biogenic silica following uptake at the roots. Factors such as solubility greatly affect the transfer of Si to the soil solution. The solubility of silicate minerals such as quartz is calculated to be 17 times lower than biogenic silica. Therefore, despite siliceous rock being the initial source of Si, the influence of these minerals on the soil

solution phase of the cycle is much smaller than the release from biogenic silica (Fraysse *et al*., 2006).

The biogenic Si cycle is the uptake and deposition of Si from the soil solution and then return though decomposition and dissolution of plant silica. The strength of the biogenic cycle influences the dynamics of Si release into the soil and is affected by the presence of vegetation and species type. For example, modelling of deciduous and coniferous forest Si cycles has revealed that the levels of Si found in above-ground biomass in deciduous forests is double the Si content found in coniferous forests. These greater levels of biogenic Si appear to influence the turnover of Si between the soil solution and vegetation, with deciduous forests calculated to have a Si uptake of 26 kg ha^{-1} yr^{-1} compared to 8 kg ha^{-1} yr^{-1} for coniferous ecosystems (Bartoli, 1983). Additionally, in the deciduous forest, biogenic silica made up 85% of soluble silicon, compared to in coniferous forests, where the biological cycle is weaker and the plant-available silicon is influenced more by the weathering of silicate minerals (Bartoli, 1983). This highlights the effect of vegetation and species type on the Si cycle between ecosystems.

The soil solution is arguably the most important reservoir of silicon for plants, as it contains the soluble form of Si, monomeric orthosilicic acid, which ranges in concentration between 0.1 mM and 0.6 mM (Epstein, 1994). Depending on the species, the soluble Si enters the plant via active or passive uptake mechanisms. At an ecosystem level, loss of Si from the cycle is caused by output of soil water or migration of dead biomass (Bartoli, 1983).

The reservoir of biological Si is produced by deposition of silica within the plant and takes the form of phytoliths or other amorphous biogenic silica structures. Phytoliths are discrete hydrated silica deposits that form through condensation and polymerisation of orthosilicic acid. The shape, size and density of these structures varies between species. For example, Mulberry (*Morus alba*) leaves contain a high density of phytoliths, with deposits ranging in diameter between 25 and 100 μm (Tsutsui *et al*., 2016).

Silicon is returned to the soil by decomposition of organic matter and dissolution of silica. In tree species this occurs over two time periods: annual or gradual litterfall; and eventual decomposition of wood tissue. This biogenic silica is insoluble and re-enters the soil solution through the

weathering processes. For example, the half-life of phytoliths from horsetail and larch in the soil matrix has been estimated to range between 10 and 12 years at pH 2 to 3 but can be reduced to less than a year at pH levels above 6 (Fraysse *et al*., 2009).

As the biogenic phase dominates the silicon cycle, the extent of plant silicon uptake is an important factor affecting the overall cycle. Species can be grouped into three classifications depending on their ability to uptake Si. These are accumulators, which deposit over 4% dry weight shoot Si, intermediates with Si content ranging between 1% and 4% and excluders that contain less than 0.5% Si (Guerriero *et al.*, 2016). Variation also occurs between higher taxonomic groups. For example, Liverworts (Marchantiophyta) and horsetail species (Equisetophyta) contain high relative shoot silicon concentrations, whereas Angiosperms, Gymnosperms and ferns (Polypodiophyta) have low shoot silicon concentrations (Hodson *et al.*, 2005).

In trees, there is significant variation in silicon uptake between species, with deciduous trees accumulating more silicon than evergreen coniferous trees. For example, the concentration of silica in the needles of Douglas fir is 1.1% dry weight, 0.97% in Norway spruce and 0.05% in black pine, compared to 1.6% in European beech leaves and 1.2% in oak (Cornelis *et al*., 2010). This study also calculated that litterfall represents the largest flux returning silicon to the soil, with 83% to 97% of silicon entering the tree being returned by this process.

Therefore, the silicon cycle is an essential process and is critical for the production of plant-available silicon in the soil solution. Input of silicon from rock minerals is important for initial access to soluble silicon, however, it becomes surpassed by the biogenic cycling of silicon between plants and soil in regions with vegetation and strong fluxes connecting these reservoirs.

1.4 Silicon accumulation in plants: uptake and deposition

1.4.1 Silicon uptake

Silicon, in the form of monomeric orthosilicic acid, may enter the plant roots from the soil solution via active and passive transport mechanisms. In species such as rice, active uptake can occur through a series of silicon

transporters and passively via mass flow of water. The soluble silicon can then be transported into the xylem and distributed to above-ground tissue by the transpiration stream.

Silicon transporters (SITs) were first identified in the marine diatom *Cylindrotheca fusiformis* (Hildebrand *et al*., 1998), where Si is required for the formation of their silicified cell walls. These *SIT* genes are present in several taxonomic groups and have been identified in some Haptophytes, Choanoflagellates and other Diatoms (Durak *et al*., 2016). Although they are not present in plant species.

The first plant silicon transporters were identified in rice (*Oryza sativa*) (Ma *et al*., 2006), which is classified as a silicon accumulator containing shoot Si concentrations ranging between approximately 2 and 5%. Therefore, silicon uptake, transport and deposition has been well characterised in rice compared to other plant species. The first plant Si transporter classified was LOW SILICON 1, which was discovered through a screen of rice mutants, identifying one defective in active silicon uptake. The screen identified healthy mutants that were unable to uptake germanium, a toxic silicon analogue (Ma *et al*., 2006). The transporter *Oryza sativa* LSI1 (OsLSI1) was subsequently identified and the family has since been expanded with the identification of a homologous gene *LSI6*.

The LSI1 and LSI6 silicon transporters are members of the Nodulin 26-like Intrinsic membrane Protein (NIP) family. These proteins are aquaporin-like channels that transport a range of solutes, including molecules such as glycerol, urea and boron (Wallace and Roberts, 2005). The structure of these proteins is crucial in determining the solutes they can transport. Each NIP protein contains a highly conserved aromatic/arginine (ar/R) selectivity filter (SF), which forms the pore of the transporter and is established by four amino acids located within helix 2, helix 5, loop E1 and loop E2 of the protein. Depending on the conserved amino acid sequence forming the SF, NIPs can be divided into three subfamilies: NIPI; NIPII; and NIPIII.

The majority of NIPI transporters contain a Trp (helix 2) - Val (helix 5) - Ala (loop E1) - Arg (loop E2) (WVAR) SF. Whereas NIPII has a more variable SF, containing the conserved arginine within the second loop, a 50% glycine/alanine probability at loop E1 and highly variable amino acids within the two helices. Finally, NIPIII has a highly conserved Gly-Ser-Gly-Arg (GSGR) SF (Trembath-Reichert *et al*., 2015).

The silicon aquaporins LSI1 and LSI6 are members of the NIPIII subfamily, therefore they contain a GSGR SF. Selectivity of silicic acid is thought to arise through two mechanisms. The first is the highly conserved SF of NIPIII proteins, where the small glycine and serine amino acids are hypothesised to create a larger pore, facilitating transport of silicic acid. In rice it has been demonstrated that the serine located in the helix 5 position of OsLSI1 is necessary for silicon specificity. Substitution mutations of this position results in complete loss of silicon uptake (Mitani-Ueno *et al*., 2011).

The second property determining silicon specificity is a precise distance of 108 amino acids between two conserved Asn-Pro-Ala (NPA) domains. Wild-type NIP2-1, a NIPIII protein, from poplar (*Populus trichocarpa*) has a distance of 108 amino acids and enables silicon uptake when transformed into *Xenopus* oocytes. However, mutating this to contain either 107 or 109 amino acids between the NPA domains results in a loss of silicon influx (Deshmukh *et al*., 2015).

In addition to the aquaporin-like Si transporters LSI1 and LSI6, another Si transporter LSI2 is involved in the transport of Si within plants. Unlike the other LSI genes, *LSI2* is predicted to encode an anion transporter with similarity to bacterial arsenite efflux protein ArsB. The LSI2 protein functions as a silicic acid/H^+ antiporter and uses active transport to drive the uptake of Si through the roots (Ma *et al*., 2007).

In rice, OsLSI1 is the initial influx transporter responsible for uptake of silicon from the soil solution. It displays specific expression in root tissue, with no expression in the rice leaf blades or leaf sheath cells (Ma *et al*., 2006). The aquaporin-like transporter has six transmembrane domains and two Asn-Pro-Ala (NPA) motifs, which are conserved structural features of LSI1 proteins. By expressing GFP under the control of the *LSI1* promoter, the authors revealed that *OsLSI1* is located in the main and lateral roots, with no expression in root hair cells. Within the rice roots, localisation of *OsLSI1* is restricted the distal plasma membrane of the exodermis and endodermis cells, found in the same region as the casparian strips (Ma *et al*., 2006), which are impermeable to solutes and restrict passive Si transport along with water uptake.

At the exodermis, OsLSI1 is required for transport of silicic acid across the casparian strip, from the epidermis into the exodermal cells. In rice, the active antiporter OsLSI2 also has polarised cellular localisation and is

positioned within the proximal membrane of the exodermal and endodermal cells. The OsLSI2 transporter located at the exodermis is able to transport silicic acid from the exodermal cells into the aerenchyma. Silicic acid then diffuses through the aerenchyma to the endodermis, where the other set of OsLSI1 transporters move silicic acid into the endodermis (Fig. 1.1). OsLSI2 at the endodermis then transports silicic acid out of these cells and silicic acid enters the xylem, where it is distributed to above-ground tissue. As Si uptake can be inhibited by treatment with metabolic inhibitors and by exposure to low temperatures (Ma *et al*., 2002), this suggests that silicon uptake in rice is through active transport, utilising ATP.

Unloading of silicic acid from the xylem is hypothesised to be carried out by the LSI1 homologue, LSI6. Unlike LSI1, this gene is expressed in the leaf xylem parenchyma cells and again shows polarised expression within these cells, with expression restricted to the plasma membrane that is proximal to the xylem vessel. The transporter appears to be responsible for distributing Si to rice leaf tissue, as *lsi6* knockout mutants contain higher concentrations of Si in the guttation fluid secreted from the xylem (Yamaji *et al*., 2008). Once transported from the xylem, the silicic acid begins to condense, and deposition occurs.

Plants can respond to changes in environmental Si by altering the expression of their LSI transporters. Transcription of *Sorghum bicolor LSI1* (*SbLSI1*) and *SbLSI2* genes are upregulated by 42.7% and 160.2% respectively, following silicon supplementation, whereas *SbLSI6* is downregulated by 33.6% (Soukup *et al*. 2017). An opposite effect is observed in the dicot soybean, where the *LSI1*-like genes *GmNIP2-1* and *GmNIP2-2* are downregulated when fertilised with potassium silicate compared to untreated conditions, displaying a decrease in root expression of 43% and 34%, respectively (Deshmukh *et al*., 2013). Despite downregulation of the soybean *LSI* genes when supplemented with Si, a 2.5 times increase in leaf Si was observed. This suggests that plants actively limit Si uptake when it is abundant in the growth medium. A similar response to environmental Si is found for *OsLSI1*, which has decreased expression following continuous supply of external Si (Ma *et al*., 2006).

Although Si accumulation tends to be high in monocot grasses, particularly in rice and silvergrass (*Miscanthus sinensis*) and primitive species such as horsetail (*Equisetum arvense*), there are an increasing number of studies

revealing the presence of Si uptake and Si transporters in dicot species. For example, soybean (*Glycine max*) is known to accumulate moderate levels of Si, containing around 0.8% leaf Si when supplemented with 1.7 mM Si from potassium silicate fertiliser. Two LSI1-like genes have been identified in this species, GmNIP2-1 and GmNIP2-2. When GmNIP2-2 is expressed in *Xenopus* oocytes, it results in Si uptake from the extracellular medium, suggesting a similar function in soybean (Deshmukh *et al*., 2013).

Homologues of OsLSI1 have been discovered in other monocot grasses, including barley (Chiba *et al*., 2009) and wheat (Montpetit *et al*., 2012). As well as in some dicots, such as soybean (Deshmukh *et al*., 2013) and pumpkin (Mitani *et al*., 2011). Differences in silicon content between a pumpkin high silicon accumulator and a low silicon cultivar was determined to be caused by a loss of CmLSI1 function in the low cultivar (Mitani *et al*., 2011). The LSI1 transporter from the Si accumulator cultivar is localised to the plasma membrane of the exodermis and endodermis when expressed in rice, matching the expression of OsLSI1. However, mutations affecting the localisation of CmLSI1 to the plasma membrane resulted in loss of silicon uptake. This illustrates the essential function of *LSI* genes in accumulating high levels of Si within plants.

Additionally, phylogenetic analysis of NIPIII genes revealed that these genes have been lost from *Arabidopsis* (Liu and Zhu, 2010). This explains the low levels of Si found in *A. thaliana*, which contains only 0.05% dry weight Si in stem tissue. Further evidence highlighting the importance of LSI1 genes in enabling Si uptake is revealed by transgenic *A. thaliana*, expressing LSI1 from wheat. These mutants result in a significant gain of stem Si, increasing from 0.05% to 0.25% (Vivancos *et al*., 2014).

There are notable differences in the root structure between rice and other species, influencing the mechanics of Si uptake. For example, aerenchyma are found in wetland and aquatic plants, including rice, however, are present to a lesser extent in other species unless exposed to hypoxic conditions. The presence of root aerenchyma reduces the frequency of cortical cells between rice exodermis and endodermis, enabling diffusion of silicic acid between these cells.

Similar to LSI2 in rice, LSI2 is expressed in the seminal and lateral roots of maize and barley. However, in these species, LSI2 is expressed only in the endodermis, where it does not display polar localisation to the proximal

membrane. Instead, the LSI2 gene can be visualised within the plasma membrane on both sides of the endodermis (Mitani *et al*., 2009). The differences in the mechanisms of Si uptake between maize, barley and rice are thought to be caused by an absence of the Casparian strip at the exodermis (Fig. 1.1).

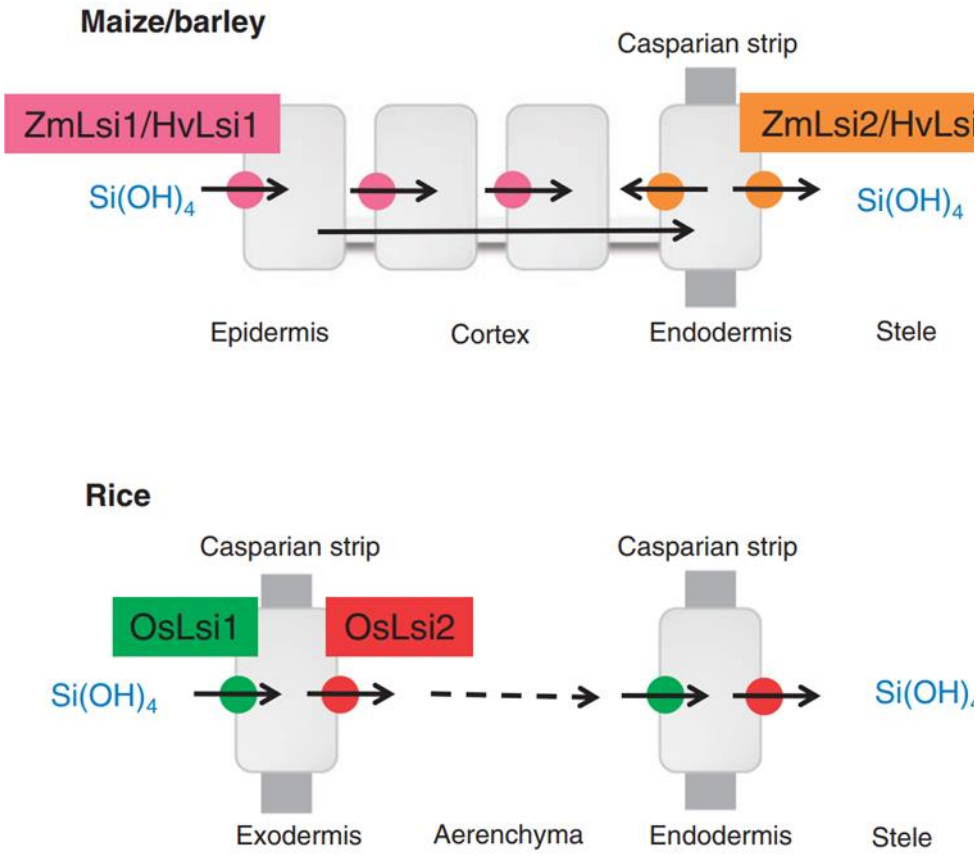

Figure 1.1. Model of silicic acid [$Si(OH)_4$] uptake from the soil solution and transport along the roots to the stele/xylem, illustrating the differences in root structure and silicon transporter (LSI) localisation between maize/barley and rice. Figure from Mitani *et al*. (2009).

Despite the progress in understanding the mechanisms of silicon uptake by the LSI aquaporins and transporters, classification of these genes is limited to a small number of monocots and even fewer dicot species. Knowledge of silicon uptake in trees is restricted to silica measurements in different tissues and silicon transporters have yet to be analysed in this group. The importance of *LSI* genes in enabling high levels of silicon accumulation is clear and the absence of these transporters relies on passive uptake along with mass flow of water, likely resulting in lower silicon deposition. Therefore, further investigation into Si uptake is critical in improving our understanding of Si in plants.

1.4.1 Silicon deposition

Silicon deposition is the condensation and polymerisation of the soluble monomeric orthosilicic acid into hydrated silica ($SiO_2.nH_2O$). This process begins with the formation of a stable nucleus of silica, which grows to form discrete amorphous silica deposits called phytoliths, or other silica structures.

The deposition process initiates when the concentration of monomeric orthosilicic acid reaches a critical level, calculated to be 100 to 200 ppm in water at 25°C (Perry and Keeling-Tucker, 2000). However, the mechanisms in which silicon reaches the critical concentration are unclear. There are two general hypotheses of silicon condensation and deposition. The first is a passive mechanism, where the concentration of silicic acid is increased as water is lost from the plant tissue by transpiration. The second is biosilicification, which involves biological molecules, including sugars, cell wall polymers and proteins, to aid the condensation of silicic acid into phytoliths or other biological silica deposits.

Arguments favouring the passive mechanism of silicon deposition are largely based on the knowledge that transportation of silicic acid occurs in the xylem, which is driven by the transpiration stream. Analysis and observations of silica in various plant tissues find that the silicon content of leaves is greater than other regions of the plant. For example, the distribution of silicon is not consistent throughout individual trees and is more concentrated in certain tissue types. Tree leaves and needles contain the largest proportion of silica, followed by the stem bark, branches and stem wood, which has low levels of deposition. In European beech trees there is 1.6% dry weight silica in leaves, 0.3% in stem bark, 0.07% in branches and only 0.004% in stem wood (Cornelis *et al*., 2010). Furthermore, in maize, the percentage dry weight of silica in the leaf blades was calculated as 10.9%, whereas, in the stem and roots the level of silicon was significantly lower, 0.8% and 2.0%, respectively (Lanning *et al*., 1980).

Additionally, in cucumber plants supplied with silicon, leaf puncture experiments revealed that damaging the leaf surface results in increased deposition around the wound site. This effect was not observed for plants grown without potassium silicate fertilisation or when plants were covered, creating high humidity conditions (Chérif *et al*., 1992). This indicates that transpiration is required for silicon deposition.

However, as silica deposits are not exclusively found in leaf tissue, the deposition process is expected to use biological molecules to initiate condensation of silicic acid and control the pattern of deposition. Heavily silicified structures have been identified in numerous grass and primitive species. The hairs of the grass *Phalaris canariensis* L. contain approximately 40% dry weight silica, horsetail branches consist of 20% silica and *Phragmites* leaves have 2% silica (Harrison, 1996). Early microscopy work by Harrison (1996), revealed that the silica structures formed in these species can be fibrous, globular or sheet-like. The presence of these structures varied based on species, with horsetail branches containing greater frequency of globular silica deposits compared to the other species. Whereas the silica found in *Phragmites* leaves varied in size considerably with no obvious structural motifs. This indicated that the molecular structure of silica can vary between different species and therefore, that biological molecules may be involved in the deposition process.

Biological silica is hypothesised to interact with plant proteins and polysaccharides. To identify the molecular interactions between silica and biological molecules, Harrison (1996) treated silica deposits with acid solutions to remove organic material. Analysis of the remaining silica extracts revealed the presence of amino acids, which were protected by silica and strongly linked to the silica deposits. The amino acids lysine and proline were present at a higher frequency in the samples treated with the acid solutions, compared to the intact leaves of *Phragmites* and the branches of horsetail. The author suggested that organic material rich in these amino acids may be involved in the initial nucleation or polymerisation of silica, based on the high level of protection they had from the acid solutions. This is consistent with the idea that silica deposits greater than 1 nm in size can begin to develop a surface negative charge, which may enable molecular interactions with positive amino acid residues or cell-wall polysaccharides (reviewed by Currie and Perry, 2007).

Additional studies have investigated the role of proteins in initiating silica deposition. *In vitro* analysis of synthetic peptides revealed that those containing a high proportion of positively charged amino acids, including lysine and arginine, were able to precipitate silica and this was proportional to the concentration of the peptide (Kauss *et al*., 2003). This provides

further evidence for the role of proteins in assisting the condensation of silicic acid.

Within plants, silica forms several different structures. These can develop as discrete phytoliths, form in the lumen of cells, hairs and trichomes or associate with the cell wall to create a silica layer. Scanning electron microscopy of silvergrass stem segments clearly illustrates this variation and reveals the presence of silica layers, silica bodies and silicified hair structures (Blecher *et al*., 2012). Based on the presence of silica located at cell walls, it is hypothesised that cell wall components, such as cellulose, callose and pectin, are involved in the formation and structuring of these silica deposits.

In horsetail plants, visualisation of silicon using PDMPO labelling revealed highly silicified stomata, discrete phytolith structures in the stems and leaves, silica-rich cell walls and potential evidence of deposition between dividing cells in the plant stem (Law and Exley, 2011). Further identification of silicon deposits in horsetail revealed the presence of a continuous silica double layer, associated with the stem epidermis, and is suspected to act as a defensive barrier to prevent fungal infection. Extensive silicification is also detected in the stomata guard cells, which could be associated with providing mechanical strength and flexibility or protect these pores from pathogen infection (Guerriero *et al*., 2018).

In vitro analysis of the effect of callose, a β-1,3-glucan and cell wall polymer, on undersaturated solutions of silicic acid revealed that the addition of callose into these solutions resulted in much larger Si particle sizes (Law and Exley, 2011). This illustrated the ability of callose to condense silicic acid, and was further illustrated by fluorescence spectrometry, which showed an increase in emission intensity for solutions treated with callose, associated with increased formation of silica structures.

Additional evidence of this is provided by analysis of *A. thaliana* mutants unable to produce and deposit callose. Mutations in the callose synthase gene PMR4 prevents the deposition of callose in the leaves and trichomes of *A. thaliana.* Loss of callose deposition resulted in a large reduction in the presence of silicified structures, especially the presence of silicified trichomes, which were abundant in wild-type and PMR4-overexpression

lines but were almost completely absent from the *pmr4* mutant (Brugiére and Exley 2017).

Further evidence towards the role of callose in silica deposition is presented in a study by Kulich *et al*. (2018). This study revealed that in *Arabidopsis* trichomes, callose appears to be required for silica deposition. It also showed that the callose synthase PMR4 and exocyst subunit 70H4, which is involved in the formation of the trichome secondary cell wall, are responsible for callose deposition in these structures. When either of these genes are mutated there is a complete loss of silicon accumulation in the trichomes.

The callose synthase mutants and callose-Si co-localisation experiments provide some evidence of this polysaccharide being involved in the formation of silica deposits in Si accumulator species such as horsetail, and also in *Arabidopsis*, a non-accumulator species. The Law and Exley (2011) *in vitro* study illustrates potential molecular interactions between silicic acid and callose, however, this has not yet been demonstrated *in vivo*.

Other cell wall polysaccharides, such as cellulose, may be involved in the ordering of silica deposits into sheet-like structures (Perry and Lu, 1992). Additionally, correlated localisations have been identified between silica and polysaccharides including arabinan-rich pectic rhamnogalacturonan in *Adiantum raddianum* (Leroux *et al*., 2013) and pectin and hemicellulose in horsetail (Gierlinger *et al*., 2008). This indicates that these cell wall polymers may be involved in silica deposition, however, causative associations have not yet been identified.

Although the process of silicon deposition has not be completely resolved, substantial progress has been made to reveal the mechanisms and molecules involved. The transpiration stream is important for the translocation of silicon within the plant and may contribute to passive deposition of silica. Additionally, biological molecules including proteins and polysaccharides could be interacting with silicic acid to promote deposition and organise silica deposits. As hypothesised in Kumar *et al*. (2017), it is also possible that separate mechanisms of silicon deposition are used in different cell types. Variation in deposition mechanisms could also occur between silicon accumulator and excluder species.

1.5 Project aims

A limited amount of information is known about levels of silicon found in tree species or the genetic mechanisms involved in uptake and deposition. Therefore, the major focus of this project is to use genome-wide techniques, including associative transcriptomics (AT) and differential gene expression analysis to identify gene markers for silicon content in ash. These techniques require less prior knowledge than reverse genetic approaches and will enable the detection of novel candidate genes. Functional analysis of the candidate markers enables hypotheses to be formed about the categories of genes associated with Si in ash trees and the biological processes involved.

These analyses are further aided by whole-genome resequencing of *F. excelsior* to improve the genetic resources available for genomic analyses, resulting in the production of a more ordered and contiguous genome and updated gene position information. Finally, the phenotype data collected for these analyses is also analysed to improve our understanding of Si uptake, distribution and accumulation in ash, providing an insight into how Si in this tree relates to other species.

Chapter 2 – Production of an improved *F. excelsior* genome

2.1 Introduction

Reference genomes provide a platform for genetic analyses and are fundamental in genome-wide association studies (GWAS) and associative transcriptomics (AT), which are powerful tools in the identification of genetic markers linked to complex phenotypes. Having a high-quality draft genome also enables comparative genomics, which aids analysis of gene and genome evolution through identification of sequence duplication, deletion, rearrangement and conservation.

Plant genomes typically contain high proportions of repetitive sequence and repeats. These regions are difficult to resolve if spanning over large regions of the genome. Therefore, a complete genome is challenging to achieve with short-read data, which are unable to cross some of these repeats, even with the use of paired-end libraries to connect contigs across repetitive regions to form scaffolds.

The current published genome assembly of *F. excelsior* was generated using an Illumina HiSeq 2000 platform and produced from short-read data. Assembly resulted in a genome composed of 89,514 nuclear scaffolds with an N_{50} of 104 kilobase pairs (kbp) and mean scaffold length of 9.7 kbp (Sollars *et al*., 2017). Based on the N_{50} statistic, the contiguity of this genome is low, meaning that it is composed of a large number of shorter scaffolds which are unordered relative to each other. The large number of scaffolds and low contiguity mean that gene position data is limited and is unable to be utilised for AT analysis. This decreases the effectiveness of the analysis, especially for single nucleotide polymorphism (SNP) data, as the gene controlling the trait of interest may not be the marker itself and instead can be located nearby in the genome.

To attempt to resolve this problem, a whole-genome resequencing project was initiated using Nanopore sequencing, with the aim of producing a more contiguous assembly. Single-molecule Nanopore sequencing provides long-read data and is dependent on the length of input DNA fragments. Lengths can reach up to 190 Kbp and averaging approximately 5.5 Kbp bases (Goodwin *et al*., 2015). These long reads have greater region of sequence consensus compared to short reads, therefore, can aid genome

assembly of highly repetitive regions. The drawback of this technology is relatively high sequencing error rates compared to other next generation sequencing methods, such as Illumina sequencing by synthesis. Nanopore 1D sequencing has been illustrated to have an accuracy of approximately 94% (Tyler *et al*., 2018), much lower than the 0.24% error rate for Illumina (Pfeiffer *et al*., 2018). However, improvements to assembly programs assist in minimising the impact of the higher error rates. For example, the Canu assembler, combined with Nanopolish (github.com/jts/nanopolish/) can produce a 98% accurate assembly of *Escherichia coli* using an input 1D dataset with only 70% accuracy (Koren *et al*., 2017).

Additionally, by re-aligning existing *F. excelsior* gene models to the more contiguous genome, the chapter also aims to improve the gene order to assist with transcriptome-based genetic analyses, such as AT.

2.2 Methods

2.2.1 DNA extraction, sequencing and assembly

To begin the genome sequencing and assembly process, frozen ash leaf samples, taken from a graft of *F. excelsior* tree 38873 (Sollars *et al*., 2017), were powdered using liquid nitrogen and a mortar and pestle, before storing at -70ºC. DNA extraction followed the manufacturer's guidelines using a CTAB and chloroform protocol (High molecular weight gDNA extraction from plant leaves protocol, Oxford Nanopore Technologies [ONT]), which utilised Carlson buffer (100 mM Tris-HCl pH 9.5, 2% CTAB, 1.4M NaCl, 1% PEG 8000, 20 mM EDTA) and further purification using QIAGEN genomic-tip, see preprint Vaillancourt and Buell (2019) for details. The extracted DNA sample was analysed with Qubit and Nanodrop prior to sequencing using PromethION and MinION technologies. Library preparation used the recommended ONT protocol, library prep SQK-LSK109 protocol, with extended length incubations and followed manufacturer's guidelines.

Canu 1.9 (Koren *et al*., 2017) was used for the assembly of sequence reads into contigs using the following parameters:

```
$ canu gridOptions="-t 48:00:00" -fast -p <assembly-directory> -d
<assembly-prefix> genomeSize=877m -nanopore-raw
<prefix>.fastq.gz
```

Canu is based on the Celera Assembler and is optimised for working with high-noise, single-molecule sequencing technologies, such as ONT (Koren *et al*., 2017). One of the initial steps in assembly is the computation of *k*-mers. These are sliding reading frames of default length 16 bases, calculated from the raw sequence reads, and are used to improve the computation of read overlaps, indicated by shared *k*-mers. To increase overlapping efficiency, Canu accounts for repetitive *k*-mer sequences by using adaptive MinHash *k*-mer weighting. In essence, this reduces the probability of a repetitive *k*-mer from being selected for overlapping, as it carries little positional information, and enriches selection of relatively unique *k*-mers.

Following overlap calculation there are multiple rounds of read and overlap correction to deal with the relatively high single-molecule sequencing error rates. To further improve assembly efficiency, Canu restricts correction to

the longest reads producing 40x coverage and uses the remaining reads to polish the assembly. Therefore, the resulting genome assembly is produced to approximately 40x coverage by default regardless of raw read coverage.

Canu is highly adapted for use on computing clusters and will automatically detect server resources and generate bash jobs to submit to schedulers such as Slurm. For this project, Canu was run using resources provided by Viking, a large Linux compute cluster available to research groups at the University of York.

2.2.2 Assembly analysis and quality control

Summary statistics of the raw sequence reads and contig sequences were generated using SeqKit version 0.12.0 (Shen *et al*., 2016). SeqKit stats function with the -a/--all flag was used to output all statistics, including sequence counts, sum length and N_{50}:

```
$ seqkit stat -a *.fasta
```

For individual contigs, properties including GC content and length were calculated using SeqKit fx2tab function with flags -n (print name only), -l (print length) and -g (print GC content):

```
$ seqkit fx2tab -n -l -g contigs.fasta
```

Whenever individual contig sequences were required they could be retrieved using SeqKit grep with flag -p/--pattern strings, which was used to identify a single contig at a time, writing the output to a text or fasta file. For multiple contigs, a text file listing contig names can be used and specified with flag -f/--pattern-file string.

```
$ seqkit grep contigs.fasta -p <contig name, e.g. tig00009581> >
contig_sequence.txt
```

To further analyse the quality of the Canu assembly, Tapestry was used to produce additional information about the contigs (github.com/johnomics/tapestry). Tapestry uses Minimap2 (Li, 2018) to realign the read set to the genome assembly, producing useful information about each contig, which was used to validate and edit the Canu assembly. These details included median read depth (MRD) of each contig, the

number and percentage of unique bases in the contig (based on alignment between contigs) and ploidy estimates that are calculated from read alignments.

The Tapestry tool is designed for use on small eukaryotic genomes. Therefore, adjustments were made to accommodate the larger ash assembly. Tapestry was run on the Viking computing cluster using approximately 70% of the read set and used a target coverage of 10x instead of the default 50x. Flags included: -c number of cores; -a assembly contigs; -r nanopore read set; -o output directory; and -d depth (genome coverage to subsample from FASTQ file).

```
$ sbatch -c 15 weave -a assembly.fasta -r reads.fastq.gz -o <output
directory> -d 10
```

As the assembly is over 50 Mbp in size, the Tapestry HTML report used to visualise and edit the contigs was not produced. Therefore, analysis of the contig_details.tsv output file was conducted using RStudio. After identifying the MRD and uniqueness values required to filter the assembly, RStudio was used to select a subset of contig names with an MRD greater than 6 or with an MRD between 1 and 6 with a uniqueness of greater than or equal to 50% (Fig. 2.1). Then SeqKit grep was used to compile a filtered genome assembly, using the "--pattern-file string" flag with the list described above.

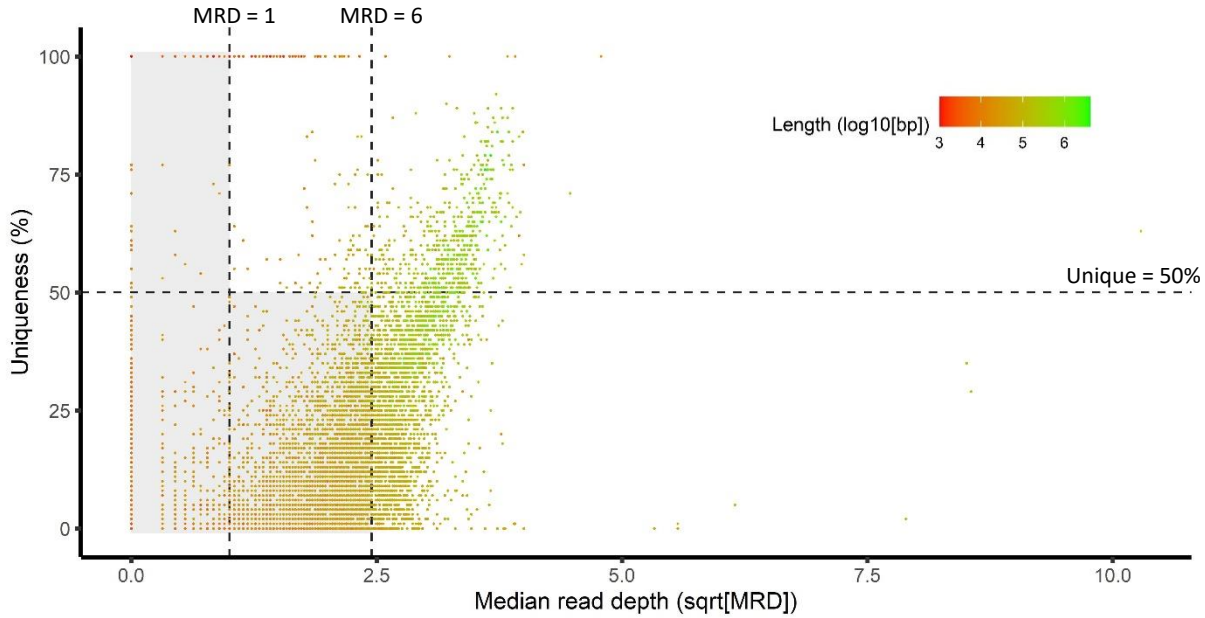

Figure 2.1. Method used to filter the genome using uniqueness and median read depth (MRD) properties. Each point represents a contig, coloured by contig length on a log10 scale. The shaded region contains the contigs filtered out of the genome after applying the filter: MRD > 6 OR (Unique >= 50% AND 1 <= MRD <= 6).

2.2.3 Identification of mitochondria and chloroplast contigs

The output dataset from Tapestry was also used to identify chloroplast and mitochondria contigs. The sequence reads for these contigs are overrepresented, therefore, the contigs contain higher MRD values. Contig sequences with a high MRD and GC content, 16 and 36% respectively, were extracted from the genome, using SeqKit, and identified as chloroplast, mitochondria or genome using BLASTn.

2.2.4 Implementing transcriptome data into the ash genome

To realign the existing transcriptome data for ash, retrieved from Ash tree genomes (ashgenome.org/data), and update gene position information, the sequence aligner GMAP (genomic mapping and alignment program; Wu and Watanabe, 2005) was used to map the transcriptome cDNA sequences to the filtered contigs. GMAP handles cDNA containing only exon sequences by splitting cDNA and genome into *k*-mers which are matched together in a process called oligomer chaining. This avoids detecting exon-exon boundaries, which can be easily influenced by sequencing errors.

Prior to mapping, the genome assembly requires processing into a GMAP database containing oligomer indexes and genomic sequence, using default *k*-mer length of 15 bases. This process was carried out using GMAP build and orders contigs alphabetically:

```
$ gmap_build --dir=<directory> --sort=alpha --db=<indexed genome name> contig.sequences.fasta
```

Following indexing of the genome, the cDNA sequences could be aligned using GMAP:

```
$ gmap --dir=<directory of indexed genome> --db=<indexed genome name> cDNA.fa > output.txt
```

This outputs a fasta-like file containing mapping positions of each cDNA to the genome, including information about multiple alignments, quality of alignment (coverage and percentage identity) and number of exons. This file was read into R using read.fasta(file, as.string = T) from the ”seqinr” package, and processed using various string manipulation functions to extract the required information.

2.3 Results

2.3.1 Production of a more contiguous *F. excelsior* genome using long-read Nanopore sequencing

The major steps involved in producing the draft assembly of ash included extraction of high molecular weight gDNA from ash leaves, sequencing using Nanopore technologies, genome assembly with Canu 1.9 and genome filtering and quality control using Tapestry and SeqKit programs.

Sequencing resulted in 16.4 million reads with total length 66,408 megabases. This is equivalent to an estimated sequence coverage of 75x, based on previous haploid genome flow cytometry measurements of 877.24 ± 1.41 megabase pairs (Mbp) (Sollars *et al*., 2017), and surpasses the minimum 20x coverage requirement for the Canu assembler (Koren *et al*., 2017). These reads were then assembled into contigs using Canu 1.9, selecting this assembler based on its optimisation for high noise, single-molecule sequencing, such as Nanopore, and its ability to utilise compute server resources for fast assembly.

Canu 1.9 produced an initial draft assembly to the default 40x coverage. Initial analysis of the genome was carried out using the data generated by SeqKit. The assembly output from Canu consisted of 8881 contig sequences, with total length of 1213 Mbp, median length of 62 Kbp and a length range of 999 bp to 4.1 Mbp. The pre-filtered assembly also had an N_{50} value of 345.8 Kbp, indicating that 50% of the genome is encoded by contigs greater than or equal to this length.

Comparisons between the expected genome size and assembly size revealed approximately 336 Mbp of excess contig sequence compared to the expected genome size of 877 Mbp. The problem could have been caused by contamination of the DNA samples prior to sequencing or a failure to collapse haplotypic regions into single contigs by the Canu assembler. This indicated that contig quality needed to be assessed and the genome filtered to a more accurate size.

GC content was the first parameter used to analyse assembly quality as it is fast to calculate and abnormal GC percentages can indicate contamination due to differences in GC prevalence between species. Based on the published data from Sollars *et al.* (2017), the average GC content of the nuclear genome is expected to be approximately 34% and

mitochondrial at 45%. This was found for the pre-filtered genome assembly, with contigs displaying GC contents of median 34.29% and mean 35.65%. GC content ranged between 1.85% and 83.77%, revealing many contigs with abnormal GC content. These sequences may be contaminants, however, as many of these occur at lower contig lengths (Fig. 2.2a), it is probable that these sequences have high repetitive GC or AT regions.

A more detailed analysis of contig quality was conducted using Tapestry, producing contig median read depth (MRD) and contig sequence uniqueness statistics, which were used to filter the initial draft assembly. The first property, MRD, is the level of coverage of each contig by the input

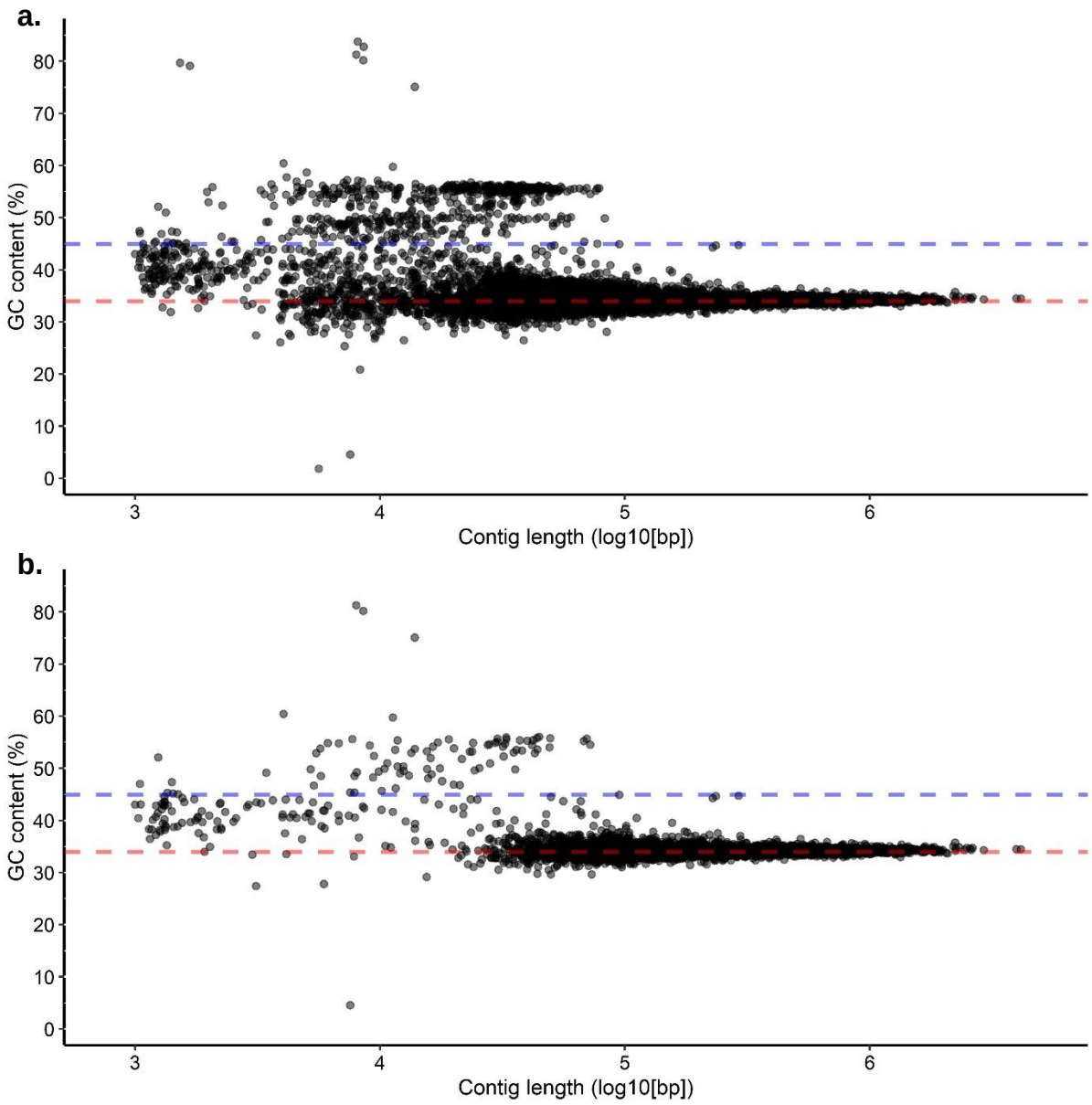

Figure 2.2. Visualisation of the **(a)** unfiltered genome and **(b)** the filtered genome. Each point is a contig, separated by length and percentage GC content. The red line is the expected nuclear genome percentage GC content (34%) and the blue line is the expected mitochondrial GC content (45%).

read set, which was sub-sampled to 10x coverage by Tapestry. A contig with an MRD of less than one means that it was unable to be fully covered by the read set. Contig sequences were filtered out of the assembly incrementally starting by removing contigs with an MRD <= 1, then <=2, <=3, etc. This was repeated until the genome reached approximately the expected size, which occurred when contigs with an MRD of less than or equal to 6 were removed from the assembly. This resulted in the removal of 5928 contigs with a sum length of 326.2 Mbp, producing an assembly of length 886.4 Mbp. This reduces the presence of contigs that were not correctly collapsed together into the haploid genome during assembly with Canu.

To prevent removal of useful sequence, the percentage uniqueness property was analysed. Uniqueness is a measure of whether a contig maps to others in the genome and was used to recover contigs with a uniqueness of greater than or equal to 50%. Out of the contigs with an MRD of less than or equal to 6, 546 have a uniqueness of greater than or equal to 50%, these contigs were recovered back into the assembly. The final stage of filtering was removal of contigs with an MRD of less than 1 to prevent the reintroduction of contaminant sequences that could be present at lower read depths (Fig. 2.1).

The filtered assembly, consisting of 3188 contigs, was reanalysed with SeqKit to produce updated summary statistics (Table 2.1). This revealed an increase in median contig length, from approximately 62 Kbp to 137 Kbp, and a mean length increase to 280 Kbp from 136 Kbp, compared to the unfiltered genome. Additionally, the filtered assembly has a total length of 893.8 Mbp, which is closer to the expected genome size, however, it is still significantly larger, suggesting that further work is required to polish the assembly. An increase in the N_{50} statistic from 346 Kbp to 568 Kbp reveals an increase in assembly contiguity, however, this is largely influenced by the decrease in assembly size.

After filtering the genome, the variation in GC content of the assembly contigs remains similar to the prefiltered genome, ranging between 4.5% and 81%. However, when visualised in figure 2.2 it is clear there has been a reduction in the number of contigs with unexpected GC contents after applying the MRD and uniqueness filters.

Comparisons to the Sollars *et al*. (2017) genome reveals that the filtered genome uses considerably fewer and longer sequences, resulting in increased genome contiguity. The filtered genome contains 3188 contigs compared to 89,514 scaffolds. It has noticeably greater sequence lengths, with the new assembly displaying 152x higher median length and 29x higher mean length. It is also significantly more contiguous than the published genome, highlighted by the 5.5x higher N_{50} value (Table 2.2). Therefore, through the use of long-read Nanopore sequencing, combined with an optimised aligner, a genome composed of fewer sequences with higher contiguity could be produced, aiding research into the genetics of ash trees.

Table 2.1. SeqKit statistics of the genome assembly pre- and post-filtered. MRD = median read depth. Lengths measured in base pairs.

	No filter	**After filtering**
Contigs	8881	3188
Sum length	1,212,658,247	893,840,713
Min length	999	999
Mean length	136,545.2	280,376.6
Max length	4,138,147	4,138,147
Median length	62,088	137,484
N_{50}	345,816	567,639

Table 2.2. Comparison of the existing published genome from Sollars *et al*. (2017) and the genome produced from this project.

	Existing genome	**New genome**
Sequences	89,514	3188
Sum length	867 Mbp	894 Mbp
Mean length	9.7 Kbp	280 Kbp
Median length	0.9 Kbp	137 Kbp
N_{50}	104 Kbp	568 Kbp

2.3.2 Identification of mitochondria and chloroplast genomes

The data generated from Tapestry was used to identify chloroplast and mitochondria contigs. The sequence reads for these contigs are overrepresented, therefore, the contigs contain higher MRD values. The GC content of these contigs was also investigated as the mitochondrial

genome is expected to contain a higher GC percentage (45%) than the nuclear genome.

The sequences of contigs with a high MRD (16) and GC (36%) content were extracted, resulting in 11 contig sequences. Their locations were predicted using BLASTn to identify the presence of nuclear, mitochondrial or chloroplast genes. Using this method, a total of four mitochondrial contigs were identified: tig00002317, tig00005526, tig00119741 and tig00119743. These had an MRD of 105.8, 28.4, 72.4 and 73.2 respectively, and a mean GC content of 44.34%. Additionally, the combined length of the mitochondrial contigs totalled 821.2 Kbp.

For the chloroplast genome, three contigs, tig0000958, tig00011941 and tig00012056, were identified as containing chloroplast genes. These contigs have a mean GC content of 37.6% and a total length of 421 Kbp.

2.3.3 Remapping cDNA sequences to improve gene position for transcriptome-based genetic analyses

To implement the new filtered genome into transcriptome-based genetic analyses, such as associative transcriptomics (AT), the existing transcriptome cDNA sequences of ash were realigned to the new genome to update gene position information. As there are fewer, yet longer contig sequences compared to the Sollars *et al.* (2017) published genome, this should increase the number of genes mapping to each contig and reduce the number of contig-contig boundaries. This should increase the ability to detect and analyse candidates from AT, especially from SNP analysis, where the gene controlling the trait of interest is not necessarily the most significantly associated SNP and can instead be located within the neighbouring genomic region, located within the SNP peak on Manhattan plots.

Alignment of cDNA transcript sequences to the filtered genome contigs was conducted using the GMAP aligner. Out of the 50,841 cDNA sequences available from Ash tree genomes, 339 (0.67%) were unable to be aligned to the contig sequences. A total of 8576 cDNAs (16.9%) mapped with multiple alignments, with a maximum of 5 alignments. However, the majority of cDNAs produced either one (82.5%) or two (16.2%) alignments (Fig. 2.3).

Additionally, GMAP predicted that 2093 cDNA sequences are possible chimeras, containing exon-exon boundaries. Although such sequences are known to occur naturally, it is also possible that these are sequencing or annotation errors within the transcriptome data.

Using GMAP to align the ash transcriptome to the genome, I was able to update the position information of 50,502 genes out of 50,841 available cDNA sequences. This aids transcriptome-based genetics analysis, and specifically enabled the creation of improved Manhattan plots following AT analysis (Chapter 4). However, the location of 8576 genes remains uncertain as these mapped to multiple regions in the genome.

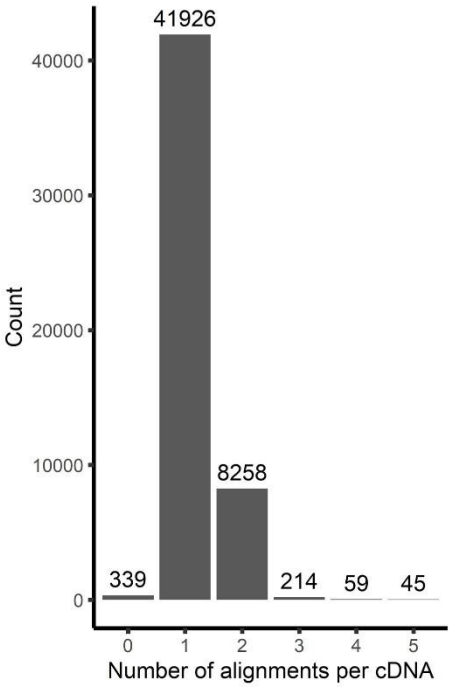

Figure 2.3. Number of alignments to the genome per cDNA, according to GMAP. A total of 50,841 cDNAs were analysed.

2.4 Discussion

The aim of this chapter was to produce an improved *F. excelsior* genome that will assist future studies investigating the genetics of ash trees. The example used in this thesis is the generation of an updated gene order, achieved through the remapping of existing transcriptome data, which will enable greater detection and analysis of SNPs from AT analyses.

The genome assembly was produced using Nanopore sequencing, the Canu assembler and a combination of data generated by SeqKit and Tapestry to filter the genome contigs. Nanopore sequencing enables the generation of long sequence reads, which are useful for spanning larger regions of the genome compared to short-read data. This assists the assembly process and allows contigs to span larger repetitive regions, found in high frequency in plant genomes. Bridging these repetitive sequences can result in improved assembly and increased contiguity, meaning the genome can be composed of fewer contig sequences. A benefit of this is improved gene order as there are fewer contig-contig boundaries.

Limitations of this sequencing technology include an increased error rate compared to other methods, such as Illumina sequencing by synthesis. To adjust for this limitation, the genome assembly was constructed using the Canu assembler, which is optimised for working with high-noise, single-molecule sequencing. It factors in the increased expected error rate and uses multiple rounds or read and overlap correction to deal with this. However, to further improve the genome quality, high-accuracy Illumina sequencing could be used to polish the Nanopore assembly. This method uses the long-read data to generate contiguous assemblies and short reads to improve overall accuracy. It is effective in producing high-quality draft assemblies of large eukaryotic genomes. For example, four *Fundulus* genomes of approximately 1.1 Gbp were assembled using a combination of Illumina and Nanopore reads. The Illumina-only assemblies used a considerably larger number of contigs compared to using Nanopore data, approximately 4.5 million contigs compared to 2,600, and displayed lower contiguity with an N_{50} of 3.7 kbp compared to 2.6 Mbp for the long-read assemblies (Johnson *et al*., 2020).

The initial genome produced by Canu was considerably larger than the haploid genome measurements from Sollars *et al*. (2017), which were

calculated by flow cytometry. Therefore, it required substantial filtering to reduce the presence of repeated contigs and produce a more accurate genome size. By using data generated from SeqKit and Tapestry programs, contig quality was assessed and contigs filtered based on their MRD and uniqueness values, which provided information about the coverage of each contig by the read set and the proportion of unique sequence compared to the other contigs in the assembly. This resulted in a genome assembly consisting of 3188 contigs and total length of 894 Mbp. More work filtering the contigs may still be required as the genome remains approximately 17 Mbp too large, indicating the filtering did not catch all the assembly errors. This may also be a result of the filtering method used, which removed contigs based on logical cut-off points.

Comparing the assembly summary statistics with the published Sollars *et al*. (2017) genome enabled the detection of improvements to the draft genome. These comparisons revealed that the assembly produced with long read data is composed of considerably fewer sequences, 3188 contigs compared to 89,514 scaffolds, and has much greater sequence lengths. The nuclear genome GC contents were found to match, with a GC percentage of 34% for both assemblies. Additionally, the N_{50} statistic was 5.5 times higher compared to the previous genome, indicating that the assembly is more contiguous. Combined, this information suggests that the assembly produced using Nanopore sequencing contains fewer contig-contig boundaries with an overall greater sequence contiguity. However, it is possible that a greater amount of sequence errors are present in the Nanopore sequenced genome, based on the higher sequencing error rate compared to the previous Illumina sequenced genome. Therefore, polishing the Nanopore data with Illumina sequencing, using software such as Pilon (Walker *et al*., 2014), should produce an assembly with high sequence contiguity and improved accuracy. For example, the genome of *Lonicera japonica* is 887 Mbp in size, similar to *F. excelsior*, and was assembled using a combination of Nanopore and Illumina technologies to produce a highly contiguous assembly with an N_{50} of 84 Mbp (Pu *et al*., 2020). Hi-C sequence data was also used in the *L. japonica* genome assembly to produce final pseudochromosome predictions. This technique involves the crosslinking of chromatin, fragmentation and ligation of interacting DNA fragments, producing Illumina sequence libraries from interacting regions of the chromatin and results in pairs of interacting DNA sequences (Belton *et*

al., 2012). The data was then used to cluster and order contigs into predicted chromosomes (Pu *et al*., 2020). Using similar approaches could further improve the quality of the *F. excelsior* draft genome.

The contig read depth and GC percentage information was also used to identify mitochondria and chloroplast contigs, detecting a total of four mitochondrial and three chloroplast sequences. The GC content of the mitochondrial contigs matches the sequences from Sollars *et al*. (2017), with GC contents of 44.34% and 44.79% respectively. However, the size of the mitochondrial genome was much larger than the previous assembly, with a total length of 821 Kbp compared to 581 Kbp. This was similar for the chloroplast contigs that have a sum length of 421 Kbp. This is larger than expected as chloroplast genomes tend to range between approximately 100 Kbp to 220 Kbp (Daniell *et al*., 2016). Therefore, it is probable that the mitochondrial and chloroplast contigs have some mis-assembly and further work is required to reveal the correct sequences. Sequence alignments between these contigs could be analysed using software such as Minimap (Li, 2018), to identify regions of overlap which would indicate assembly errors.

The ability to map the ash transcriptome to a genome with fewer contig-contig boundaries enables improved ordering of genes, as there are more genes located within each contig. Out of the 50,841 available cDNA sequences, the majority (82.5%) were mapped to a single location in the genome. However, 8576 cDNAs mapped with multiple alignments, therefore, could not be narrowed down to a single locus and their position remains uncertain. Additionally, there were 339 unmapped cDNA sequences, which could be located in the contigs removed from the assembly or remained unmapped due to an absence of a significant alignment. The presence of around 2000 cDNAs marked as chimeric with exon-exon boundaries by GMAP could indicate the presence of sequencing or annotation errors within the transcriptome data. Therefore, reannotation may be required in the future to further improve the genomic and transcriptomic resources available for *F. excelsior*. One way to achieve this would be to align existing RNA-sequence reads to the improved genome, using software such as HISAT2 (Kim *et al*., 2019), producing a new collection of cDNA and coding sequence (CDS) models. The aim of this would be to create a more accurate CDS model and correct the existing

chimeric cDNA sequences. Alternatively, further RNA sequencing could be conducted, increasing the read depth available for transcriptome assembly.

Overall, the use of long-read sequencing technologies to produce a draft genome of ash has reduced the number of contig sequences and increased genome contiguity. Remapping the transcriptome has improved gene order and resulted in a greater number of genes per contig sequence. This should improve the detection, analysis and plotting of SNP and gene expression marker (GEM) candidates from transcriptome-based genetic analyses such as AT.

It is apparent that further work is required to polish the current assembly and additional sequencing is necessary before the number of contigs can be considerably reduced and the genome resolved to a chromosome level. As Nanopore sequencing was used, future studies could also map and then investigate the methylation data, providing insight into the epigenome/methylome of ash.

Chapter 3 – Physiology of silicon in ash

3.1 Introduction

The initial step in understanding the biology of Si in ash trees is to determine the extent of Si uptake and accumulation, which have yet to be identified in ash leaf and branch tissue. The degree of Si accumulation varies between plant species which can be classified as Si accumulators, excluders and intermediate types. A difference between plants can be observed at higher phylogenetic levels. In general, Angiosperms contain greater amounts of shoot Si compared to Gymnosperms and ferns (Polypodiophyta), and less than some species of moss, horsetail and liverworts (Hodson *et al*., 2005). However, classification into higher taxonomic groups is inaccurate as substantial variation in Si accumulation occurs within groups, at a species level. Even in known Si accumulators with *LSI* genes, there is considerable variation in Si deposition. For example, cucumber plants express *LSI* genes and accumulate approximately 1% dry weight shoot Si. Whereas the shoot Si content of rice cultivars can vary between 1.5% to 4% (Talukdar *et al*., 2019), and in some cases have been found to accumulate up to 5% silica (Ma *et al*., 2006).

Additionally, Si accumulation in tree species is inadequately understood. Trees such as mulberry are observed to contain a high density of Si phytoliths (Tsutsui *et al*., 2016), whereas other species have intermediate levels of Si, such as European beech containing 1.6% dry weight leaf silica, or particularly low amounts of Si, for example, 0.05% Si for black pine tree needles (Cornelis *et al*., 2010).

Therefore, understanding the basic physiology of Si uptake, distribution and deposition are fundamental in improving our knowledge of the mechanisms of these processes. To achieve this, the chapter analyses leaf Si data collected from 20 *F. excelsior* trees at two separate time points approximately a month apart. This data is used to reveal the extent of Si accumulation in leaf tissue and investigate the process of Si uptake in ash. Additionally, it analyses several datasets collected in other sections of this thesis, including branch Si data and ash dieback (ADB) disease severity scores, recorded from a diverse group of 125 ash accessions (Chapter 4), as well as using previous leaf Si measurements collected from the same

set of individuals (Turnbull and Harper, unpublished) to determine the distribution of Si within the tree.

The overall aim of this section is to improve our understanding of Si uptake, distribution and association with the ADB disease. It also introduces the leaf Si data used for differential gene expression analysis (Chapter 5) and the branch Si data later analysed using associative transcriptomics (Chapter 4).

3.2 Methods

3.2.1 Ash leaf sampling and silicon analysis

20 ash trees of similar age, approximately 10 to 15 years old, were selected for sampling from a field site at The Hawkhills, York (54.101° N, -1.178° W). Half of these were fertilised with 0.5 mL potassium silicate fertiliser in 2 L of H_2O on 12/07/2019. Two sets of leaf samples were collected from each tree on 12/07/2019 and on 19/08/2019, after 33 days. One set of samples were then dried overnight at 65^{o}C and another stored at -80^{o}C, after snap freezing in liquid nitrogen, for RNA-Seq differential gene expression analysis (see Chapter 5).

To prepare the leaves for X-ray fluorescence (XRF) analysis, the dried leaf samples were powdered using a Retsch MM400 ball mill, set at a frequency of 20 beats s^{-1} for three minutes, using two 10 mm metal balls. The powder was then formed into a maximum of three 13 mm diameter pellets, fewer if sufficient material was not available. Pellet formation used a manual hydraulic Specac press at ~10 kg x 1000 tons and a 13 mm die. There was no attempt to control the thickness of the pellet as this does not affect the XRF silicon readings (Reidinger *et al*., 2012). To prevent transfer of powder between samples, the die was cleaned after each sample and the ball milling equipment washed and dried between uses.

The Si content of the leaf samples was measured using an XRF analyser (Reidinger *et al*., 2012). Each pellet was run once on each side for 20 seconds, using up to three pellets for each tree. These technical replicates account for variation in pellet formation and XRF analysis. Measurements were carried out in a low-grade helium atmosphere to increase detection of Si.

To standardise Si values and ensure the XRF analyser was calibrated, two samples of known Si concentration were run every 10 pellets. Si measurements were normalised to the calibration samples, accounting for non-biological variation in Si readings by the analyser. This involved calculating an overall average of all calibration sample readings. Then, for each set of calibration readings, every 10 samples, a normalising coefficient was calculated, which is the local calibration average divided by the overall calibration average. This normalising coefficient was used to modify the Si readings by dividing the raw Si reading by the normalising

coefficient. Si change is calculated using August:July ratio and percentage change.

The methods of sampling and measuring branch Si and ADB disease severity are found in Chapter 4. Additionally, comparisons between leaf and branch Si contents were conducted using leaf Si data produced by Lucy Turnbull in the Harper Lab. This data was collected from 103 of the same ash accessions as the branch Si panel, using the XRF method to measure Si content with a single replicate for each individual tree (Turnbull and Harper, unpublished). These leaf samples were collected in July 2019, earlier than the branch samples.

3.3 Results

3.3.1 Ash tree leaves accumulate silicon over time

There is currently no literature describing the levels of Si found in *F. excelsior*, the distribution of Si in different tissues or whether ash trees accumulate silicon over time. Therefore, to answer these questions, a series of experiments were carried out that involved measuring leaf Si at different time points and recording branch Si levels. To achieve this, a set of 20 ash trees were selected for leaf sampling at two time points, once on the 12th of July 2019 and again on the 19th of August 2019. Half of these trees were fertilised with a potassium silicate solution to determine the effect of Si fertilisation on Si accumulation.

The leaf samples collected before fertilisation in July show a mean Si level of 0.21% and variation ranging between 0.09% and 0.31%. After a month, Si content increased in 18 of 20 trees, resulting in a mean Si level of 0.26% for fertilised trees and 0.30% for unfertilised trees. Variation in Si is also observed in the August samples, ranging from 0.15% to 0.37% in fertilised trees and 0.21% to 0.42% in unfertilised trees (Fig. 3.1a). The difference between fertilisation treatment groups is not significant (Wilcoxon rank sum:

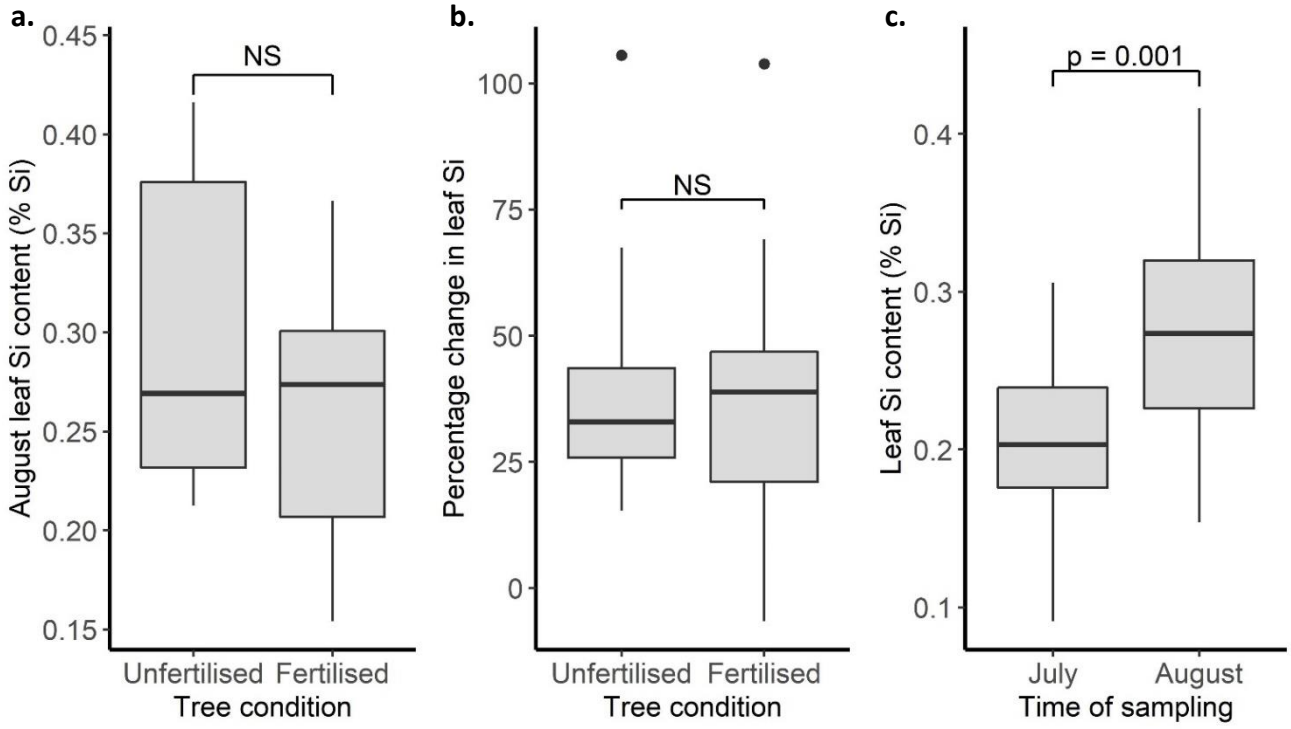

Figure 3.1. Analysis of leaf silicon data. **(a)** Leaf silicon content of samples collected in August from trees either fertilised with potassium silicate or left unfertilised (n=10 per group). **(b)** percentage change in leaf silicon between samples collected in July and August. **(c)** leaf Si measurements from July and August samples.

W = 61, p = 0.44), indicating that fertilisation using 0.5 mL potassium silicate fertiliser in 2 L of H_2O has no effect on leaf silicon accumulation within a month after treatment.

There is substantial variation in the ability of individual trees to accumulate Si over time. Percentage increase in Si varied within fertilised and unfertilised groups, ranging from -6.60% to 103.85% in fertilised trees and 15.32% to 105.58% in unfertilised trees. The mean increase in leaf Si was by 39% for fertilised trees and 41% for unfertilised trees. Again, fertilising with potassium silicate had no significant effect on percentage change of leaf silicon (Wilcoxon rank sum: W = 47; p = 0.8534) (Fig. 3.1b).

As fertilisation had no effect on leaf Si, samples were ungrouped and the difference between July and August samples was tested. There was a significant increase in leaf Si content over the month (t-test: t = 3.445; d.f. = 38; p = 0.001), with a mean gain of approximately 0.07% Si (Fig. 3.1c). These results reveal that ash can uptake, transport and deposit Si in the leaf, resulting in accumulation over time. Whereas, the absence of an effect of Si fertilisation on leaf Si content suggests that either soil Si is not a limiting factor in Si deposition, or that the quantity or time of fertilisation was not sufficient for an effect to be observed.

3.3.2 Branch tissue accumulates significantly less Si than leaves

To investigate the distribution of Si throughout ash trees, Si measurements were recorded from 125 branch samples and compared with leaf Si data (Turnbull and Harper, unpublished). Analysis of the branch Si data revealed that across the 125 trees, branch Si ranges between 0.024% and 0.077% with a mean of 0.045% (Fig. 3.2a). Comparison with leaf Si data revealed there are significantly higher Si levels in the leaves compared to branches (Wilcoxon rank sum: W = 0, p < 0.001) (Fig. 3.2b).

Additionally, calculation of a leaf to branch ratio indicates that silicon tends to be 3.57x higher in leaves (median) and ranges between 2x and 6x (Fig. 3.2c). This is consistent with existing literature of Si in tree species, which show that SiO_2 content is always higher in leaf than branch tissue (Cornelis *et al.*, 2010), despite a different Si measuring method being used.

While processing branch samples for XRF Si analysis, twig diameter measurements were collected to account for variation in branch thickness. Comparisons between Si content and branch diameter reveal that there is a

weak but significant, negative association between these properties (Linear regression: F = 111.2; d.f. = 1,372; p < 0.001; R^2 = 0.23), with lower percentage Si content present in thicker branches (Fig. 3.3). Given that Si measurements are collected as a percentage, this effect could be a result of decreasing branch surface area as diameter increases, which affects the proportions of bark and wood in the sample.

In addition to Si data, ash dieback disease severity scores were recorded at the field site in November 2019 (see Chapter 4 for method). This measurement considers the approximate percentage of the tree displaying visible ash dieback symptoms, including cankers, lesions and branch discolouration, with a score of 5 representing the severe infected trees and 1 for the least symptomatic trees. This data was collected to provide an indication about whether Si plays a role in resistance to ash dieback. Analysis of this data revealed that there is a significant association between ash dieback disease severity and branch Si relative to diameter (% Si mm^{-1}) values (Kruskal-Wallis: chi-squared = 31.596, d.f. = 4, p < 0.001). A post-hoc test highlighted that the trees with the most severe infection, a disease score of 5, have significantly lower branch Si than trees with a less severe disease score. No significance was detected between the other disease score values (Fig. 3.4). The association between ash dieback disease severity and branch Si, indicates a link between these factors.

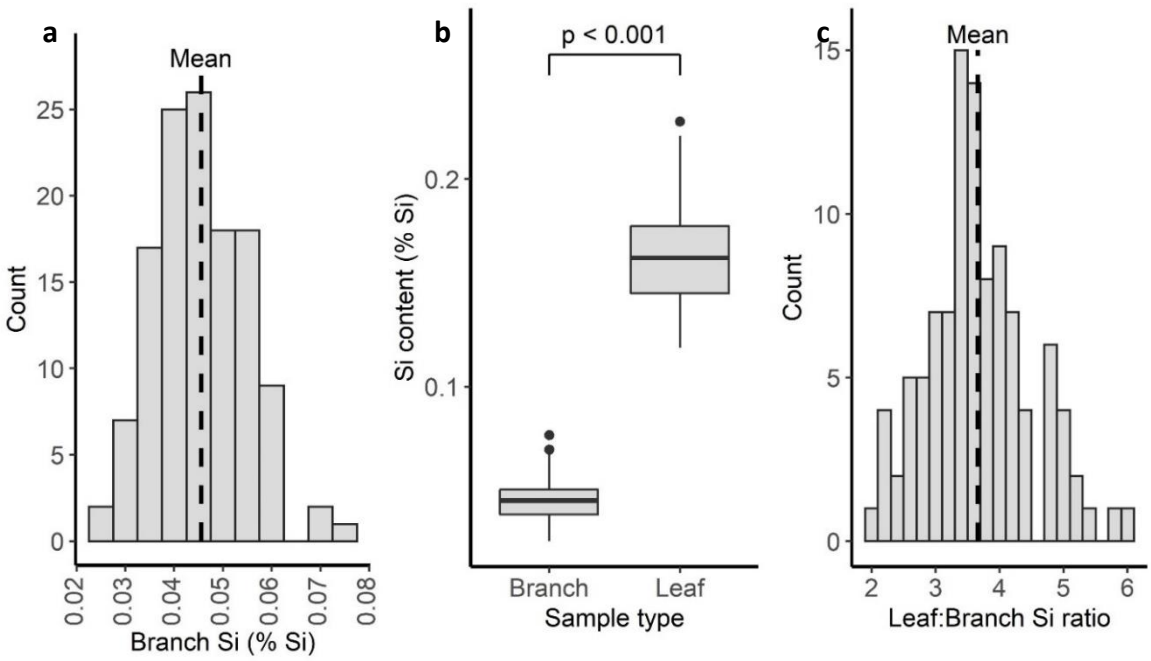

Figure 3.2. Analysis of branch and leaf Si data. **(a)** Variation in branch Si content across 125 ash trees. **(b)** Comparison between branch and leaf Si levels. **(c)** Leaf to branch Si ratio (n = 103). Each accession, n, is a mean of three biological replicates.

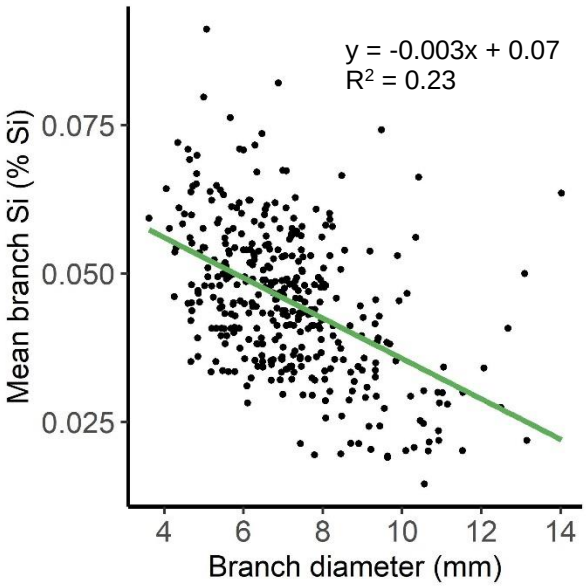

Figure 3.3. Linear regression between branch diameter and branch Si content. n = 374. Plot annotated with regression line equation and R^2 value.

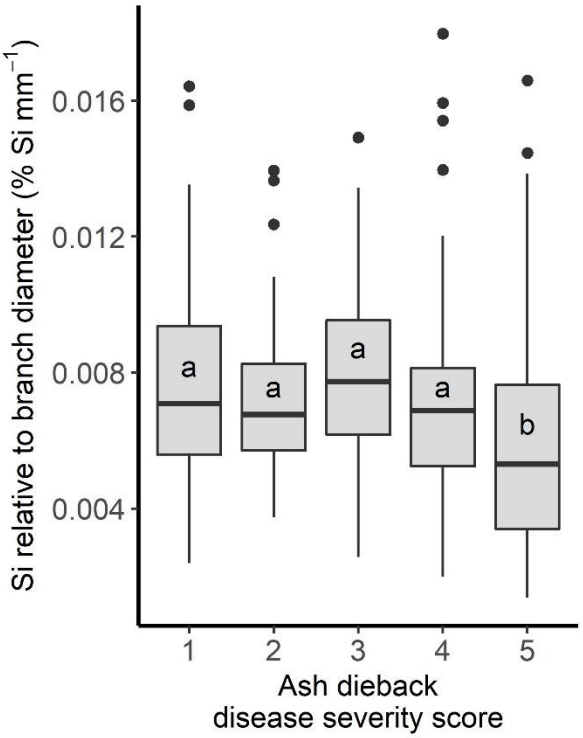

Figure 3.4. Branch Si content found in ash trees with a range of ash dieback disease severity scores. Boxplots annotated with a different letter are significantly different. Sample counts are 88, 70, 68, 72 and 121, for disease scores 1 to 5, respectively.

3.4 Discussion

The aim of this chapter was to improve our understanding of Si uptake and accumulation in ash trees. To achieve this, the Si content of leaf and branch tissue were measured using XRF analysis, enabling the investigation of Si fertilisation on Si levels, Si uptake, distribution of Si within the tree and association with ash dieback infection.

Leaf Si content was calculated to be an average 0.21% in July and 0.28% in August, which was a significant increase of approximately 0.07%. This demonstrates an ability of ash to accumulate low amounts of leaf Si over a relatively short period and is a similar to observations made in two citrus species, *Citrus sinensis* and *Citrus reticulata*, which display significantly higher silicon levels in mature leaves compared to young leaves (Mvondo-She and Marais, 2019). However, variation in percentage increase of leaf Si was detected, ranging from a decrease of 6.6% to an increase of 106%, and did not accumulate for two of the twenty trees. This could be explained by low sample size as only a single leaf was collected per tree. Therefore, the Si measurements may have been influenced by factors such as distance from roots and maturity of the leaf.

Based on the low level of Si within the leaf tissue, it appears that ash does not actively accumulate Si. Active accumulator species have high Si concentrations, for example, horsetail branches have a dry weight silica content of 20%, *Phalaris canariensis* hairs contain 40% dry weight silica, *Phragmites* leaves have 2% silica (Harrison, 1996), rice stems contain 2% to 5% Si (Ma *et al*., 2006) and barley has 1.6% leaf Si (Grašič *et al*., 2019). Whereas, non-accumulators contain lower Si concentrations, such as *A. thaliana* that has a dry weight stem Si content of 0.05% (Vivancos *et al*., 2014).

Analysis of branch samples found that the Si content of this tissue was on average 0.045%. Comparing this with previous leaf Si measurements of the same trees, demonstrated that the Si content of leaf tissue is always greater than the tree branches, ranging between 2x and 6x higher. This is consistent with previous studies revealing that leaf or needle Si content is always higher than the branch Si concentration. Cornelis *et al*. (2010) found this occurs in Douglas fir, Norway spruce, black pine, European beech and Oak. Although the difference between leaf and branch Si is greater than ash for most of these species, except for black pine. For example, the leaf

and branch dry weight silica content of European beech was 1.59% and 0.07%, respectively, which is 22.7x higher in the leaves. Whereas black pine has a similar leaf to branch ratio to ash with a 1.6x higher needle Si content. This suggests there is considerable variation in the accumulation of Si between tree species.

Comparing branch Si values with twig diameter data revealed a weak negative association between these factors. As Si content is calculated as a percentage, it is possible that a surface area effect is being observed and Si is not evenly distributed throughout the branch, with greater levels in the bark rather than wood tissue. A previous study investigating the distribution of Si throughout trees have revealed that bark tissue contains a higher proportion of Si compared to the stem wood. For example, in European beech, stem bark has 0.32% dry weight silica, whereas the stem wood contains only 0.004% silica (Cornelis *et al*., 2010). If a similar effect occurs in ash, then branches with a high diameter will contain proportionally less bark and more wood tissue compared to thin branches, and therefore they may have a lower percentage Si content. A future study to confirm this hypothesis could separate bark and wood tissue from ash branches and measure the Si content in these tissues to determine location of Si within the branch. If the hypothesis is correct, then a higher Si content should be found in the bark tissue.

The Si fertilisation experiment was conducted to test the effect of environmental Si on Si accumulation in ash. No association between Si fertilisation and leaf Si content was observed, which suggests that either environmental Si was not the limiting factor for Si uptake and deposition, or that the quantity or duration of fertilisation was not large enough for an effect to be observed. Further experiments could analyse different quantities of Si fertiliser or designed using tree grafts and controlled silicon levels in the soil, rather than using field samples with varying soil conditions.

The analysis of branch Si and ash dieback disease severity revealed that trees with the most severe ADB symptoms contained significantly less branch Si, however, no differences were observed between the other disease scores, which could be explained by a difficulty to assign trees accurately to scores 2 to 4. A wide variety of research indicates the beneficial effects of Si in increasing disease resistance to fungal

pathogens, including in olive trees (Nascimento-Silva *et al*., 2019), rice (Nakata *et al*., 2008) and banana plants (Kablan *et al*., 2012). Therefore, it is possible that Si contributes to resistance to ash dieback and prevents the most severe infection symptoms. However, as ADB results in crown dieback and branch tissue necrosis, Si uptake, transport and deposition are likely compromised during ash dieback infection. Additionally, as the field site is naturally infected, individual trees at the site have been affected by the pathogen for different periods and are at various stages of infection. This will influence the reliability of the disease score measurement.

Further work will need to be completed to answer the ADB hypotheses. Studies should be conducted to investigate the association between ADB infection and branch Si levels and overcome some of the limitation of this experiment caused by the field layout and uncontrolled Si levels in the soil. ADB inoculation experiments of ash in controlled conditions with altered concentrations of Si fertilisation would provide an improved method of determining whether Si contributes to resistance to ADB or just instead impedes Si uptake through disease damage.

This investigation provides an insight into the general levels of Si found within ash tree. It reveals that ash only accumulates low levels of Si, favouring deposition in leaf tissue compared to its branches. It finds that leaf Si increases as the leaves mature and identifies a possible association between ADB and branch Si. As the study only reveals the general concentrations of Si in ash, future work could focus on analysing the distribution of Si in more detail, through analysis of bark and wood samples, or investigate the form of Si deposits. Microscopy could be used to identify the structure of Si deposition in ash leaves and bark through visualisation of discrete phytolith structures or cell wall-associated Si, providing additional information about the biological role of Si within ash trees.

Chapter 4 – Associative transcriptomics analysis of branch silicon data

4.1 Introduction

Silicon is a non-essential plant nutrient that is known to improve disease resistance to a diverse range of pathogens, insects and other herbivores (see Chapter 1). However, there is limited information on the genetics controlling silicon uptake and deposition in plants, or about the impact of silicon on gene expression. We know that uptake can be driven by a group of aquaporin-like transporters, termed LSI1 and LSI6, and by a Si efflux transporter LSI2. Previous research also indicates the role of proteins and polysaccharides in initiating and organising Si deposition, for example, callose and lysine-rich proteins are thought to able to cause the condensation of soluble Si into silica deposits (Law and Exley, 2011; Harrison, 1996).

Transcriptome-based studies, such as Jiang *et al*. (2019), have revealed that Si may cause differential expression of several defence-associated and signalling genes. Si can also influence the levels of key defence hormones including ethylene, jasmonic acid and salicylic acid, potentially explaining the link between Si and disease resistance.

As the genetics associated with Si in plants remains reasonably unknown, a genome-wide approach was taken, which does not rely on as much prior knowledge as reverse genetics approaches. Associative transcriptomics (AT) was used to identify genetic markers associated with variation branch Si content across a diversity panel. These markers can be SNPs, signifying sequence changes linked to the input trait data, or GEMs, which explain variation in the trait of interest based on the expression level of individual genes.

Therefore, this chapter aims to investigate the genetics associated with Si deposition in ash, revealing genetic markers linked to branch Si content. It uses a diverse panel of 114 British and Northern Irish *F. excelsior* accessions to provide the phenotypic and genetic diversity required for AT. Functional analysis of these markers may provide insight into the processes controlling Si accumulation or reveal genes regulated by biogenic silica.

4.2 Methods

4.2.1 Field site and branch sample collection

Permission was granted to sample branch material from the Cemetery Fields Forestry Research site (55.855° N, 3.19° W) on the 25th and 26th of November 2019. The site contains 276 *Fraxinus excelsior* archive accessions collected from across the United Kingdom, which were systematically placed in 28 columns with a maximum of three replicate grafts per accession. A limited number of other tree species are also present and ash dieback infection was noted for the majority of ash trees in the field. The *F. excelsior* diversity panel used for the AT analysis consists of 114 of the 276 accessions, all of which have three replicate grafts.

Where possible, twig samples were collected from upright branches and regions clear of visible ash dieback infection. Additionally, ash dieback disease severity scores were recorded for each tree based on the approximate percentage of the tree infected with the fungus (Table 4.1). As twig sampling occurred after leaf abscission, ash dieback disease severity was judged based on visible branch infection, including cankers, lesions and branch discolouration, resulting from dead tissue (Fig. 4.1).

Table 4.1. Ash dieback disease severity score key

Score	Percentage of tree infected
1	0 - 20
2	21 - 40
3	41 - 60
4	61 - 80
5	81 - 100

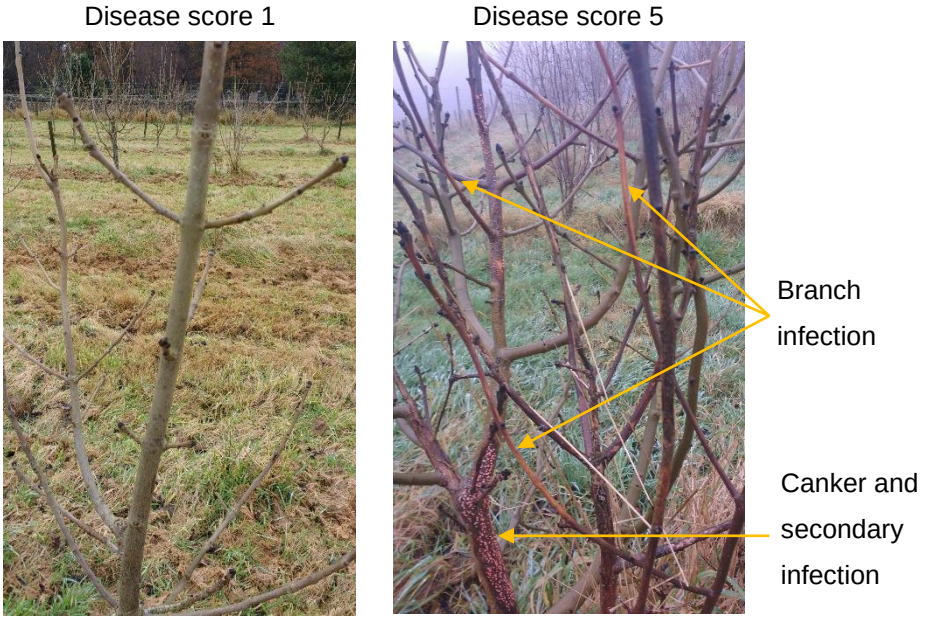

Figure 4.1. Field site images of trees with an ash dieback disease severity score of 1 and 5.

4.2.2 Branch sample processing and XRF silicon analysis

To prepare the branch samples for XRF analysis, they were first oven dried for 48 hours at 65^{o}C, followed by powdering using a ball mill. A similar method was used for this step as Chapter 3, except a single 20 mm ball was used instead of 2 x 10 mm balls, to deal with the tougher material. Prior to ball milling, branch samples were processed into smaller pieces using secateurs, to assist the powdering process. Additionally, twig diameter measurements were collected using digital callipers with a 0.01 mm resolution and +/- 0.02 mm accuracy. The powder was then processed into pellets using a manual press (see Chapter 3).

XRF silicon measurements were collected using the same protocol as the leaf samples (Chapter 3). However, the calibration samples were re-run every 12 XRF measurements instead of 10, to keep the three replicates of each accession together and improve sample tracking, and no additional technical replicates were measured. The calibration readings were again used to correct the Si measurements and Si reads were divided by the twig diameter value to account for variation in twig size, resulting in the measurement branch Si content relative to diameter (% Si mm^{-1}).

4.2.3 Associative transcriptomics of branch Si traits

An associative transcriptomics approach was used to investigate the genetics of silicon deposition in ash and identify genetic markers linked to silicon content. AT is a type of genome-wide association scan (GWAS) that uses mRNA sequences instead of genomic DNA to identify genes associated with variation in a trait of interest. To achieve this, AT combines population and phenotype data, collected from a diverse panel of plant accessions, with genetic markers, which can be SNPs or GEMs. The AT analysis of trait data is based on similar methods of previous studies (Miller *et al*., 2016; Havlickova *et al*., 2017) and sequencing for this panel was completed in 2016, prior to this AT analysis. This analysis was repeated for multiple variations of the branch Si trait, including branch Si content and Si relative to branch diameter, as well as for the ash dieback disease severity score data.

GAPIT was used to identify associations between SNPs and the phenotype traits (Lipka *et al*., 2012). This is achieved with mixed linear models, inputting 694,887 SNPs and incorporating population structure estimates,

which are calculated with PSIKO (Popescu *et al*., 2014), and population relatedness data from a VanRaden kinship matrix. Estimates of population structure and relatedness are necessary to reduce the likelihood of type I false positive errors. This is because population structure affects linkage disequilibrium, defined as the tendency for markers to associate with traits. This becomes more susceptible to error in closely related populations as these groups have similar markers that are not necessarily associated with a causative gene to the input trait data.

Furthermore, GEMs are calculated using a fixed-effect linear model that uses RPKM gene expression data and PSIKO population structure estimates as explanatory variables and the phenotype data as the response variable. RPKM data is filtered to remove genes with expression less than 0.4 RPKM. The GEM analysis used gene expression data from 40,133 genes, of which 7197 had an RPKM of less than 0.4.

To further reduce the likelihood of selecting false positive candidates from the SNP and GEM analyses, two parameters were used, false discovery rate (FDR; Benjamini and Hochberg, 1995) and the Bonferroni multiple test correction threshold. The Bonferroni threshold is a stringent adjustment of the 0.05 significance threshold, calculated by dividing 0.05 by the number of markers in the analysis. Whereas FDR is a less-stringent modification of the *p* value and is calculated using the equation:

> FDR = p value x Number of markers / p value rank position (ascending)

To visualise the output of the AT analyses, Manhattan plots were produced that show the genome position of the markers against the significance of association to the trait data ($-\log_{10}[p]$). Marker position information was retrieved by aligning the existing transcriptome sequences to the filtered draft genome presented in Chapter 2. The genome requires further sequencing and assembly to resolve chromosomes, meaning that genome contigs were ordered alphabetically in the Manhattan plots. However, gene position within contigs is predicted and where gene position is uncertain, these are represented with a different symbol or marked as “NA” for unknown position. These plots were produced using ggplot2 in R version 3.6.0 (R Core Team, 2017).

4.2.4 Bioinformatic analysis of genetic markers

Further analysis of GEMs and SNPs was conducted using a range of online bioinformatic tools and resources. NCBI BLASTx (Altschul *et al*., 1997) and TAIR were used to identify related genes and proteins. Detection of conserved domains was conducted using the NCBI conserved domain database (Lu *et al.,* 2020). GEM protein sequences were predicted using ExPASy translate tool (web.expasy.org/translate). EMBOSS Needle pairwise sequence alignment tool was used to analyse similarity between two protein sequences (ebi.ac.uk/Tools/psa/emboss_needle).

4.3 Results

4.3.1 Associative transcriptomics reveals genetic markers associated with branch Si content

The aim of conducting the AT analysis was to get an insight into the possible genetic mechanisms associated with Si levels and processes such as Si uptake and deposition. To achieve this, the AT analysis processes the input trait data and outputs a list of GEMs or SNPs that associate with the branch Si input traits, pass the FDR and exceed the more stringent Bonferroni multiple test correction threshold.

No significant SNPs were retrieved from the GAPIT analysis of the branch Si and Si relative to diameter traits. Therefore, no further analyses were carried out on this data. However, analysis of the branch Si relative to diameter trait resulted in 647 GEMs passing the 0.05 FDR threshold, with nine of these exceeding the Bonferroni threshold (Table 4.2; Fig. 4.2). A single significant GEM was detected for the unadjusted branch Si trait, passing both the Bonferroni threshold and FDR significance level (Table 4.3; Fig. 4.3). This GEM is FRAEX38873_v2_000344050.1, which is also the third most significant GEM from the Si relative to diameter trait analysis. Another GEM, FRAEX38873_v2_000172740.1, was also detected from the analysis of both traits, however, its association to the branch Si trait is less significant than to the Si relative to diameter trait, as the FDR does not pass the 0.05 threshold (FDR = 0.065). This GEM is predicted to encode a poly(A) polymerase (Table 4.2).

Table 4.2. GEMs associated with the Si relative to branch diameter trait, passing FDR and Bonferroni thresholds.

Gene ID	P value	Trait correlation	BLASTX annotation
FRAEX38873_v2_000172740.1	1.27E-07	Positive	Poly(A) polymerase 1-like
FRAEX38873_v2_000362370.1	1.61E-07	Negative	Probable serine/threonine-protein kinase
FRAEX38873_v2_000344050.1	1.69E-07	Negative	Peptidyl-prolyl cis-trans isomerase 1
FRAEX38873_v2_000278950.1	2.14E-07	Positive	Cullin-4-like
FRAEX38873_v2_000387490.1	5.29E-07	Positive	Proline-rich receptor-like protein kinase
FRAEX38873_v2_000387350.1	6.23E-07	Positive	Uncharacterised / chromatin-associated
FRAEX38873_v2_000011420.1	6.30E-07	Positive	Probable serine/threonine-protein kinase / IREH1-like
FRAEX38873_v2_000198760.1	7.29E-07	Positive	Cullin-4-like
FRAEX38873_v2_000046980.1	1.44E-06	Positive	Uncharacterised / apoptotic chromatin condensation inducer in the nucleus

Table 4.3. GEM associated with the branch Si content (%) trait, passing FDR and Bonferroni thresholds.

Gene ID	P value	Trait correlation	BLASTX annotation
FRAEX38873_v2_000344050.1	2.44E-07	Negative	Peptidyl-prolyl cis-trans isomerase 1

The GEMs were further analysed using bioinformatic tools and by exploring the current literature. Out of the nine GEMs, five display potential links to silicon or are associated with processes that may affect or be affected by silicon levels. These are: a peptidyl-prolyl cis-trans isomerase (PPIase - FRAEX38873_v2_000344050.1); two cullin-4-like proteins (CUL4-1/2 - FRAEX38873_v2_000278950.1 and FRAEX38873_v2_000198760.1); a proline-rich extensin-like receptor kinase (FRAEX38873_v2_000387490.1); and a serine/threonine protein kinase IREH1-like (FRAEX38873_v2_000011420.1).

The other four GEMs are predicted to encode uncharacterised proteins or have non-specific roles. These include: a poly(A) polymerase 1 protein (FRAEX38873_v2_000172740.1); a serine/threonine protein kinase (FRAEX38873_v2_000362370.1); an uncharacterised protein (FRAEX38873_v2_000387350.1) containing two agenet domains of unknown function (smart00743; pfam05641) and similarity to a gene of unknown function in *A. thaliana* (AT4G17330); and an uncharacterised, apoptosis-related protein (FRAEX38873_v2_000046980.1), which consists of several DNA/RNA interaction domains including an RNA recognition motif (cd12432), a splicing-associated protein binding motif (pfam16294) and a putative DNA binding domain (smart00513).

One of the most significant GEMs, encoding a PPIase, shows high similarity to a cyclophilin peptidyl-prolyl cis-trans isomerase from *Olea europaea* (BLASTx: 95% cover; 96.43% identity; XP_022896740.1), a closely related species to *F. excelsior*, as well as several cyclophilins including AtCYP18-2, AtCYP18-4, AtCYP19-1, AtCYP19-2, AtCYP19-3 and AtCYP20-2, when the search is restricted to *Arabidopsis* proteins. This GEM displays a significant negative correlation with branch silicon content (Linear regression: $F = 31.43$; $d.f. = 1, 112$; $p < 0.001$), with high marker expression associating with low silicon content (Fig. 4.4a). The protein

encoded by this GEM contains a cyclophilin-type PPIase domain between amino acids (AAs) 4 and 84. This protein family is highly diverse; however, members are often involved in protein folding.

In comparison to the PPIase, the GEM FRAEX38873_v2_000387490.1 has a positive association with branch silicon content (Linear regression: F = 28.62; *d.f.* = 1, 112; p < 0.001; Fig. 4.4b) and a BLASTx search reveals sequence similarity to proline-rich receptor-like protein kinase, PERK9, from *Olea europaea* (BLASTx: 58% cover; 80% identity; XP_022878386.1). Restricting the search to *Arabidopsis thaliana* results in top hits to AtPERK8, 9, 10 and 13, indicating that this GEM is a member of the PERK family. An interleukin-1 receptor associated kinase (IRAK), serine/threonine catalytic domain is located in the predicted protein sequence between AAs 363 and 628. This domain is present in key plant developmental and signalling proteins, including CLAVATA1, WUSCHEL and BAK1.

There are two *CUL4* genes positively associated with branch silicon content (Fig. 4.4c, d). These are homologues with high sequence similarity (EMBOSS Needle: 93.7% identity; 96.4% similarity; 0.5% gaps) and display moderately correlated expression patterns (Linear regression: F = 285.8; *d.f.* = 1, 245; p < 0.001; R^2 = 0.54), potentially indicating co-expression.

FRAEX38873_v2_000278950.1, referred to in this chapter as *CUL4-1*, is most similar to a cullin-4-like protein from *O. europaea* (BLASTx: 95% cover; 96.02% identity; XP_022848611.1). Whereas FRAEX38873_v2_000198760.1 (*CUL4-2*) is more closely matched to a cullin-4-like isoform (BLASTx: 99% cover; 95.15% identity; XP_022898841.1). Both GEMs BLAST to *A. thaliana* CUL4 (AT5G46210.1) when limited to *Arabidopsis* proteins.

Analysis of conserved domains revealed that both ash CUL4 proteins have predicted cullin family sequences between AAs 146-743 for CUL4-1 and AAs 149-745 for CUL4-2. As well as a conserved cullin neddylation domain between AAs 771-835 or 773-837, which is involved in the binding of NEDD8 proteins.

The final GEM, FRAEX38873_v2_000011420.1, is predicted to encode the serine/threonine protein kinase IREH1 based on sequence similarity to an *Olea europaea* homologue (BLASTx: 99% cover; 97% identity). This gene also has high similarity to two *Arabidopsis* genes *INCOMPLETE ROOT HAIR ELONGATION 1* and *3, IREH1* (AT3G17850) and *IRE3* (AT1G48490)

respectively. The GEM displays a positive association with branch Si content (Linear regression: $F = 28.1$; $d.f. = 1, 112$; $p < 0.001$), with greater expression in high Si trees (Fig. 4.4e). Its predicted protein structure consists of 525 AAs with a microtubule-associated serine/threonine-like (MAST-like) catalytic domain spanning AAs 117-405.

The identification of GEMs associated with the Si relative to diameter and branch Si traits, reveals potential genetic control of Si accumulation or regulation of gene pathways by Si. The majority of these GEMs have predicted functions in growth and developmental processes, indicating that these influence the uptake or deposition of Si in ash trees.

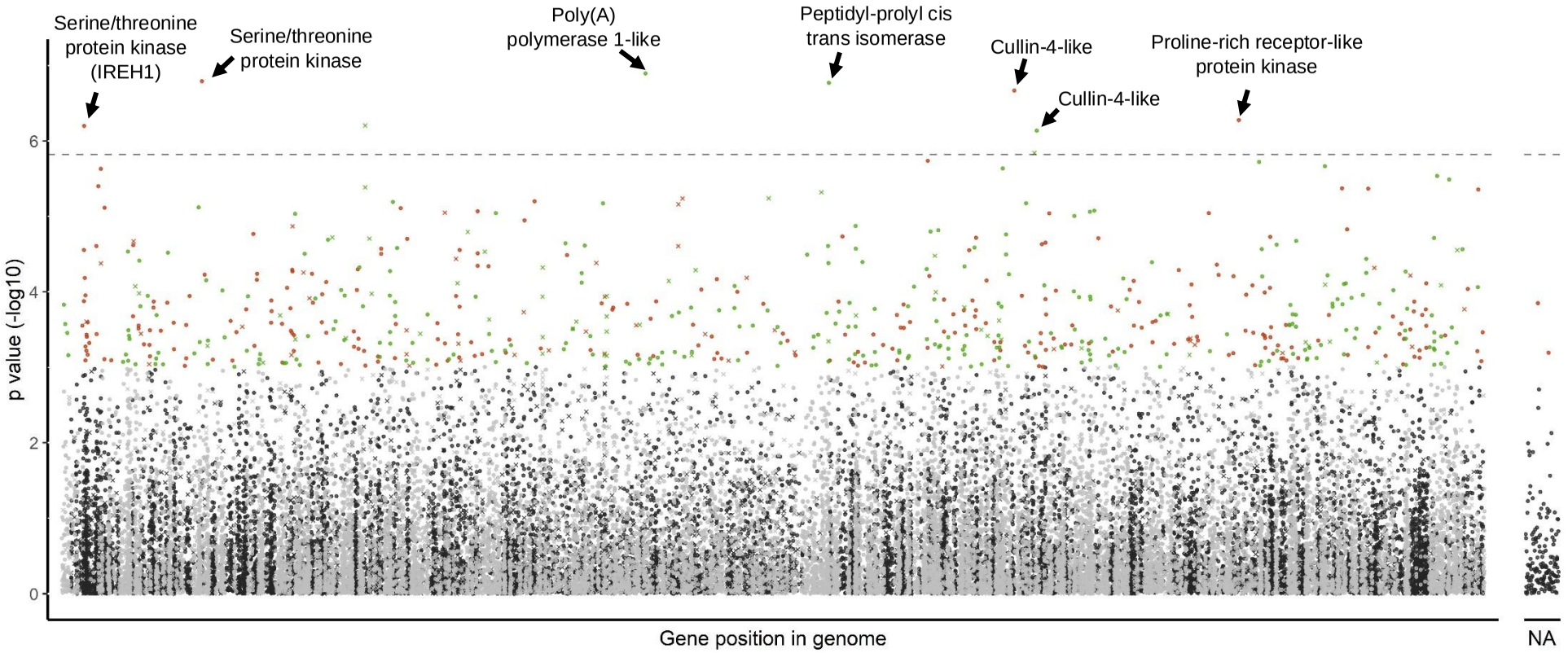

Figure 4.2. Transcript expression associated with the Si relative to branch diameter trait. The significance of trait association is displayed as a -$\log_{10}$(p) scale. Each point represents a gene and are pseudo-positioned based on their location in the genome, predicted in Chapter 2. Genome contigs are ordered alphabetically and are identified by alternating colours of grey/black for genes that do not pass FDR and green/orange for GEMs that pass FDR. Within contigs, genes are positioned based on their sequence mapping. Genes with multiple alignments are represented as symbol (**x**), and single mappings with (●). In the case of multiple alignments, only a single position is plotted. Genes with no significant alignments to the genome are located in the "NA" panel. The dashed line represents the Bonferroni multiple test correction threshold. The GEMs discussed in this chapter are labelled with their BLASTX annotation.

63

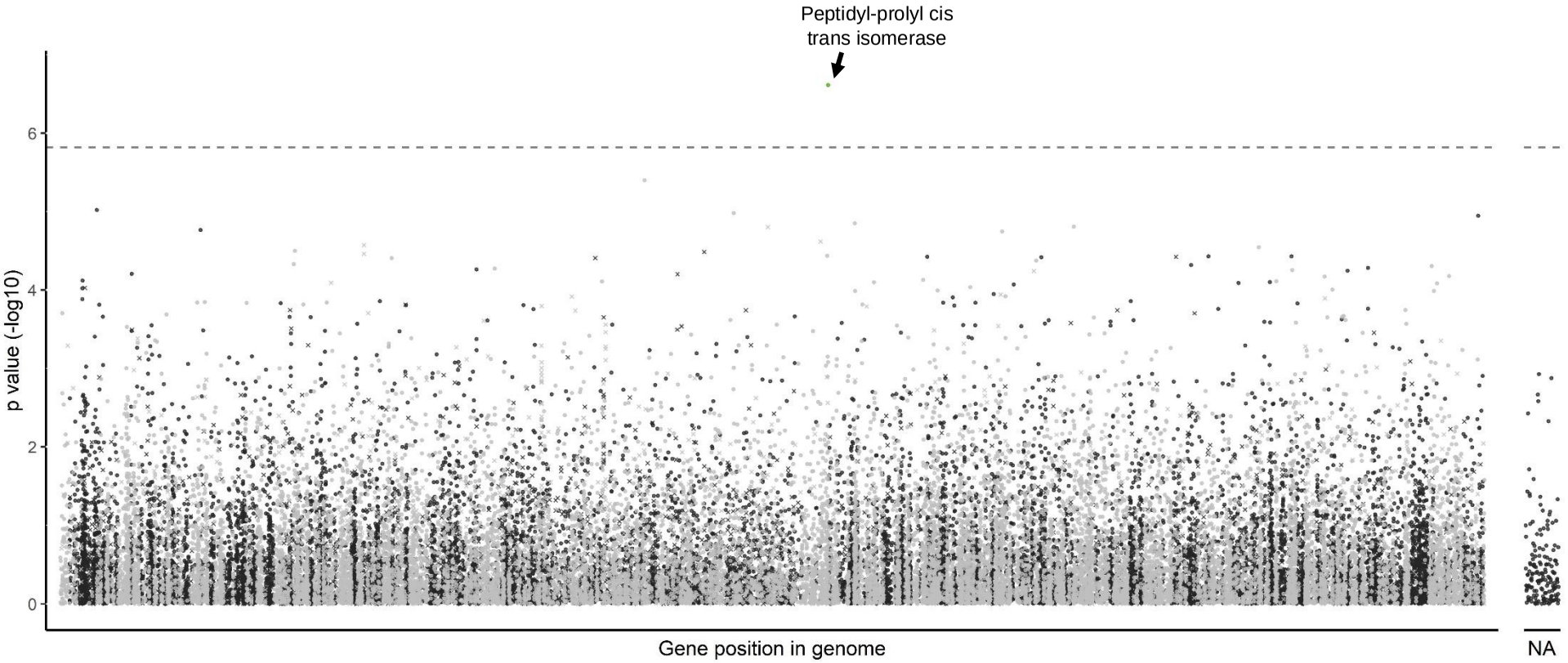

Figure 4.3. Transcript expression associated with the branch Si trait. The significance of trait association is displayed as a -$\log_{10}$(p) scale. Each point represents a gene and are pseudo-positioned based on their location in the genome, predicted in Chapter 2. Genome contigs are ordered alphabetically and are identified by alternating colours of grey/black for genes that do not pass FDR and green/orange for GEMs that pass FDR. Within contigs, genes are positioned based on their sequence mapping. Genes with multiple alignments are represented as symbol (**x**), and single mappings with (●). In the case of multiple alignments, only a single position is plotted. Genes with no significant alignments to the genome are located in the "NA" panel. The dashed line represents the Bonferroni multiple test correction threshold. The GEMs discussed in this chapter are labelled with their BLASTX annotation

64

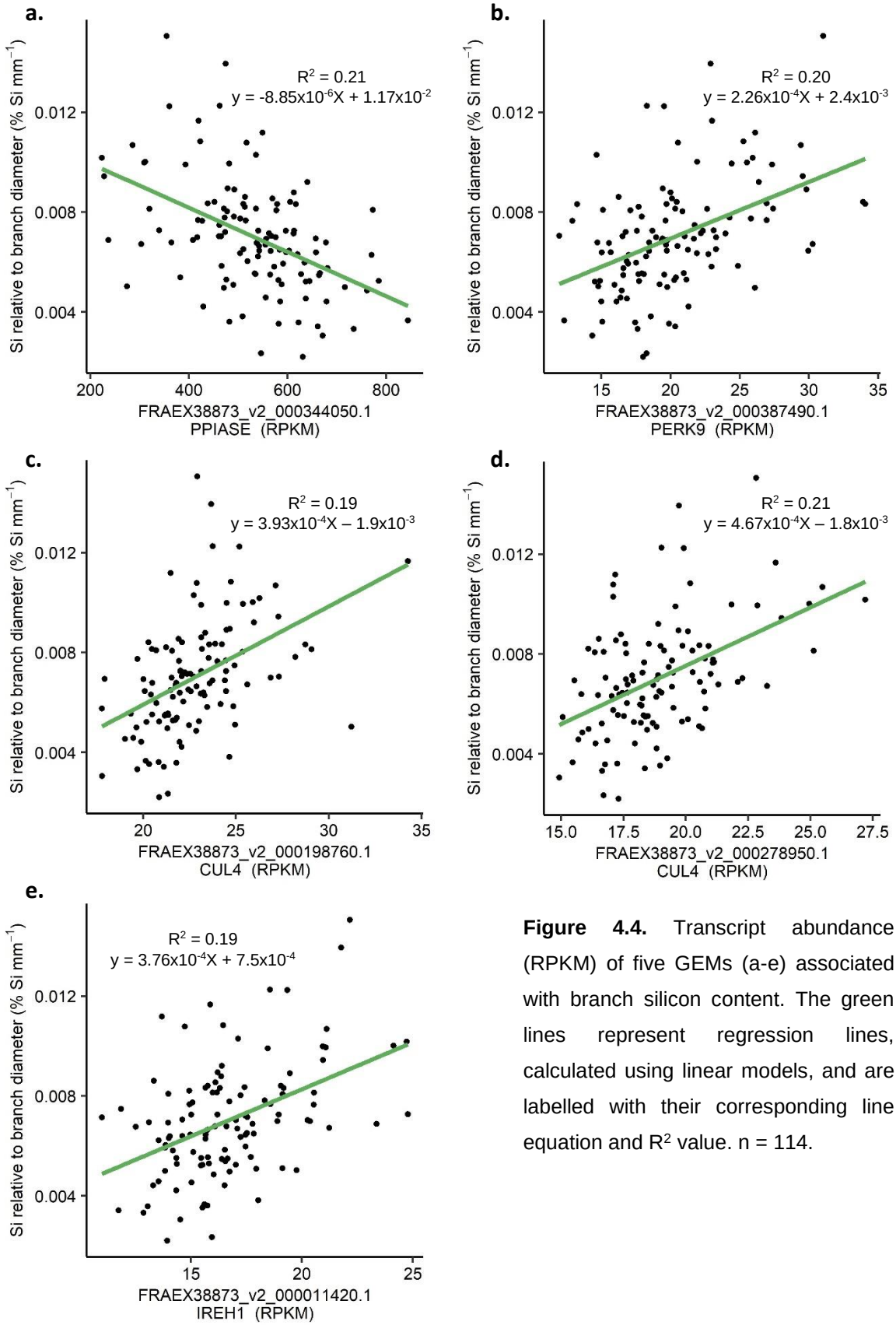

Figure 4.4. Transcript abundance (RPKM) of five GEMs (a-e) associated with branch silicon content. The green lines represent regression lines, calculated using linear models, and are labelled with their corresponding line equation and R^2 value. n = 114.

4.4 Discussion

The overall aim of this chapter was to identify genetic markers linked to branch Si content in ash trees, using the AT approach. This was beneficial as it does not require prior biological knowledge of the trait of interest and in this case, enabled the discovery of a selection of GEMs that associate with the concentration of Si in branch tissue.

The AT analysis resulted in the identification of nine GEMs with highly significant association to the Si relative to diameter phenotype trait. One of the most significant GEMs outputted from the analysis of both branch Si and Si relative to diameter traits is predicted to encode a cyclophilin PPIase. The PPIase gene family is extensive and ubiquitous across both prokaryotes and eukaryotes and is divided into three classes based on protein structure: Cyclosporin A (CsA)-binding cyclophilins; FK506-binding proteins (FKBPs); and Parvulin-like PPIases (Shaw, 2002).

The main role of cyclophilin PPIases is to aid the folding of proteins by catalysing the *cis-trans* isomerism of peptide bonds of proline residues. They are suspected to be involved in a wide array of biotic, abiotic and developmental processes (Romano *et al*., 2004), regulating these through targeted folding of key proteins in signalling and regulatory pathways. There are a predicted 35 duplicated cyclophilin genes in *Arabidopsis*, which have varying expression patterns throughout leaf, flower, root and stem tissue (Romano *et al*., 2004). These factors indicate a diversification of function and a role in a range of biological processes.

An example of regulation by cyclophilins is the control of auxin signalling by *Lycopersicon esculentum* CYP1 (LeCYP1). Treatment of hypocotyl with auxin (1-NAA) revealed a reduced sensitivity to this phytohormone by *dgt* mutants compared to wild type (Oh *et al*., 2006). This was illustrated by reduced overall rooting by the *dgt* mutants at a variety of auxin concentrations. The *dgt* allele contains mutations in *LeCYP1*, resulting in the loss of protein expression of this cyclophilin. The *dgt* phenotype can be partially replicated by application of CsA to wild-type plants, which inhibits PPIase activity. A notable phenotype of the *dgt* mutation is the complete lack of lateral root formation when cultured in phytohormone-free medium. Therefore, revealing that cyclophilins are involved in key processes, such as auxin signalling, and are crucial for lateral root development.

Cyclophilin PPIases have also been associated with the phytohormone gibberellic acid. In wheat, a dwarf mutant *gaid* is insensitive to gibberellic acid. Application of this phytohormone resulted in increases in the length of the first leaf sheath and blade, as well as increasing the first leaf sheath cell length. However, in *gaid* mutants, limited differences were observed between plants with or without gibberellic acid treatment, indicating reduced sensitivity (Li *et al*., 2010). The dwarf mutant was found to express TaCYP20-2 at a higher level than wild-type plants. Additionally, a TaCYP20-2 overexpression line resulted in a similar reduced height phenotype, producing plants that are approximately 10cm shorter. This suggests that TaCYP20-2 has a negative effect on stem elongation and may promote insensitivity to gibberellic acid.

A study investigating wild-type ROC1 (AtCYP18-2) and a mutant *roc1* with a gain-of-function point mutation, revealed novel molecular details of ROC1 in *A. thaliana* (Ma *et al*., 2013). The authors discovered that overexpression of ROC1 using a CaMV 35S promoter and RNAi knockdown of ROC1 resulted in no clear plant architecture phenotypes. However, a single point mutation from C to T at nucleotide position 173 was sufficient to alter the amino acid sequence, converting a serine to phenylalanine. This small sequence change generated a mutant phenotype (*roc1/+*) displaying significantly altered plant architecture, with considerably shorter stems, 50% truncation in stem epidermal cells, a 3-times increase in stem branching and altered leaf morphology with shorter petioles. This reveals the potential for closely related PPIases to ROC1 to have roles in key leaf and stem developmental processes.

Additionally, localisation of *ROC1* transcripts by expressing ROC1::GUS revealed that ROC1 is expressed in elongating tissue, including root tips, leaf petioles and emerging stems (Ma *et al*., 2013). As well as in the vascular tissue of cotyledons and rosette leaves. The expression pattern of ROC1 therefore provides further evidence towards the role of ROC1 in growth and development.

There is some evidence linking Si to PPIase expression. In rice, production of a mutant overexpressing the Si uptake aquaporin LSI1 resulted in significant increase in silicon content (Fang *et al*., 2017). This LSI overexpression mutant has reduced expression of a PPIase gene compared to wild-type rice plants.

Despite the functions of the majority of cyclophilin PPIases being unknown, several have now been revealed to be involved in plant developmental processes and interact with key phytohormones. Therefore, it is possible that the GEM is also involved in processes such as root, stem, leaf or vascular tissue formation. These processes may influence Si content for several reasons: initially root growth may affect water availability and access to soluble Si in the soil; leaf development will affect transpiration and so may alter Si transport and deposition; and vascular tissue formation could influence transport of Si from the roots to apical tissue. As the GEM has a negative association with branch Si content, it is possible that the PPIase is acting as a negative regulator of these developmental processes and therefore restricting silicon uptake at high expression levels.

Two of the GEMs positively associated with branch Si content are predicted to encode CUL4 proteins. The CUL4 protein is a scaffolding subunit of the E3 ubiquitin ligase complex, which targets proteins for degradation by the 26S proteasome, controlling gene expression and regulating developmental processes. In *Arabidopsis*, CUL4 is expressed constitutively but also shows specific higher expression in leaf vascular tissue, mature primary roots, lateral roots and pollen, illustrated through expression of GUS under the control of the *AtCUL4* promoter (Bernhardt *et al*., 2006).

Compared to wild-type plants, knockdown of *AtCUL4* has a significant impact on leaf development, displaying reduced overall growth, irregular leaf shape, reduced leaf number and underdeveloped vascular tissue, which are left unconnected to the main vascular system. These mutants also have abnormal root phenotypes, containing fewer lateral roots that wild-type plants (Bernhardt *et al*., 2006). Therefore, CUL4 appears to be critical in enabling the correct development of leaf, vascular and root tissue.

As loss of *CUL4* expression in *A*. *thaliana* has substantial negative effects on these developmental processes, the *CUL4* GEMs in ash could impact branch Si uptake by enabling correct root formation. Alternativity, transpiration, a factor hypothesised to be involved in Si transport and deposition, would be influenced by correct leaf and vascular tissue development.

Out of the nine significant GEMs, two were predicted to encode proteins with kinase activity. These were PERK9-like and IREH1-like proteins and may suggest a role of signalling networks in Si accumulation or regulation

of these networks by Si within the plant. In *Arabidopsis,* PERKs are membrane-bound receptors that are predicted to be involved in sensing the environment at the cell wall and are thought to have roles in sensing cell wall expansion and integrity. AtPERK8, 9 and 10 are able to interact with other signalling molecules, such as AGC kinases class VIIIa, including AGC1-9 and kinesin-like calmodulin-binding protein (KCBP)-interacting protein kinase (KIPK), demonstrated through yeast two-hybrid screens using the cytosolic domain of AtPERK10 and partial clones of PERK8 and 9 (Humphrey *et al*., 2015). This illustrates the role these proteins have in interactions with other signalling molecules, which are predicted to be involved in growth signalling pathways (Bögre *et al*., 2003).

The study by Humphrey *et al.* (2015) predicted that these genes contribute to the negative regulation of root growth after triple *perk8,9,10* mutants displayed slightly increased (1.1x) root length in the presence of high sucrose. Conversely, overexpression of PERK10 using a DEX-inducible promoter resulted in the arrest of root growth in some of the transgenic seedlings. This would in theory limit Si uptake when highly expressed by restricting root growth, however, there was no change in root length under control conditions.

Multiple PERK proteins appear to influence cell wall composition. The PERK10 overexpression mutants displayed formation of ectopic lignin and callose in the primary root and hypocotyl. This was also illustrated in transgenic *Arabidopsis* expressing PERK1 from *Brassica napus*, which appeared to show ectopic deposition of callose and cellulose in the roots of these transgenic plants through histochemical staining (Haffani *et al*., 2006). Therefore, high expression of FRAEX38873_v2_000387490.1 may affect the levels of callose in the cell wall. This could influence Si deposition as callose has been shown to cause the condensation of undersaturated solutions of silicic acid *in vitro* (Law and Exley, 2011). Additionally, loss of callose deposition through mutations in the callose synthase PMR4, result in loss of silicified trichomes (Brugiére and Exley, 2017), providing some evidence towards the role of callose in Si accumulation.

Another GEM with kinase activity was predicted to encode the INCOMPLETE ROOT HAIR ELONGATION protein, an AGC “other” class of serine/threonine kinase. As some AGC kinases are known to interact with AtPERK9 proteins, it is possible that the IREH1-like and PERK9-like

GEMs are interacting to influence Si content, although interaction experiments would have to be carried out to investigate this.

In *A. thaliana, IREH1* and *IRE3* genes are interacting proteins and are known to have roles in root development and architecture (Yue *et al*., 2019). Knockout mutations of *Arabidopsis ireh1-1/-2* displayed root skewing phenotypes, resulting from rotation in root epidermal cells. A similar phenotype was shown by *ire3* knockouts and a double mutant enhanced the skewing further.

One of the ways that these genes appear to influence root development is by enabling correct formation and alignment of microtubules in the developing root tissue. Microtubule arrangement has also been visualised in the *ireh1, ire3* and double mutant, by crossing the microtubule marker *KNpro*:EGFP-MBD with the mutant lines. This showed a reduction in the stability and density of microtubules in the root meristem, elongation and maturation zones of the mutants. Additionally, in the maturation zone of the mutant plant roots, there was significant misalignment of microtubules, with over 30% of the microtubule arranged in different directions, whereas in wild-type roots, microtubules were perpendicular to the direction of growth (Yue *et al*., 2019).

Furthermore, *ire Arabidopsis* mutants display approximately 40% reduction in root hair length, demonstrating the role this protein kinase has in promoting root hair elongation (Oyama *et al*., 2002). Therefore, high expression may contribute to correctly formed root structure and promotes silicon uptake by enabling sufficient water uptake, explaining the positive correlation observed between FRAEX38873_v2_000011420.1 expression and silicon content.

The identification of GEMs associated with the Si relative to diameter and branch Si traits, reveals potential genetic control of Si accumulation or regulation of gene pathways by Si. The majority of these GEMs have predicted functions in growth and development of leaf, root and vascular tissue, indicating that these influence the uptake or deposition of Si in ash trees. More specifically, it appears that proteins involved in cell elongation, cytoskeletal development and controlling cell wall composition, are the main genetic factors associated with Si levels in ash trees.

No significant SNPs were detected for both branch Si traits. This could be a result of environmental factors influencing Si content or caused by

limitations of the AT analysis. It is possible that Si accumulation is a complex trait, controlled by multiple additive loci, which individually have weaker effects on Si content, rather than by fewer strong effect loci that are easier to detect. The GEM analysis illustrated that growth and developmental processes are associated with branch Si content. As such processes involve the majority of expressed genes, it could explain why no significant SNPs were detected.

Additionally, the low number of accessions in the AT diversity panel reduces the power of the analysis by limiting trait and genetic variation available in the population. However, the genotypes used in this study are wild accessions or have limited breeding selection, so sufficient variation should be available.

Furthermore, the AT analysis uses transcript data collected from ash leaf tissue. This can restrict the genetic analysis by preventing identification of genes that are inducible or have specific expression patterns in other tissue types in the tree. This could have prevented the identification of significant associations for root specific genes that may affect uptake of silicic acid from the soil solution. An example of this are the *LSI1* and *LSI2* genes which are known to have very specific expression and are restricted to cells in the root endodermis and exodermis (Ma *et al*., 2011).

Alternatively, environmental factors may contribute to variation in silicon content. There are considerable limitations and issues with the field site used to sample ash branch material. The 276 accessions present at the field site contain up to three replicate grafts that were planted systematically down 28 columns. Therefore, replicates were not randomised creating the potential for field effects to influence the Si data. Factors that influence the branch Si data could include soil composition, Si availability, terrain gradient and time infected with ash dieback, which would have varied between trees as the field has become naturally infected with the fungus.

Chapter 5 – Differential gene expression analysis of leaf Si data

5.1 Introduction

There is a limited understanding about the genetics behind Si deposition in leaf tissue or the effects Si has on regulating gene expression. Several differential gene expression analyses have been previously conducted to examine the effects of Si on transcription profiles under a variety of condition, including fungal and bacterial pathogen inoculation, salt stress, drought stress and metal toxicity (reviewed by Manivannan and Ahn, 2017). These studies have found that Si increases the expression of photosynthetic components, including subunits of photosystem I and II (Song *et al*., 2014), which is thought to assist plants in dealing with stress.

Additionally, Si appears able to upregulate the expression of defence-associated genes when exposed to bacterial or fungal stresses, such as pathogenesis-related proteins, chitinase, peroxidase enzymes and stress-responsive factors (reviewed by Manivannan and Ahn, 2017). For example, when Si treated tomato plants were exposed to the fungal pathogen *R. solanacearum*, genes including PRRs, PR genes, SAR-related genes and stress-related genes, were upregulated compared to untreated controls (Jiang *et al*., 2019). Modulation of hormone signalling by Si was also observed, with significant differences in the production of jasmonic acid, ethylene and salicylic acid between Si and non-Si groups, following inoculation with the pathogen. This reveals that Si may have the ability to regulate gene expression of critical biotic and abiotic stress-responsive genes. However, the role of Si in altering the gene expression of tree species, including ash, has yet to be investigated.

Therefore, the chapter uses differential gene expression analysis to identify genes with significant expression differences between five ash trees with high leaf Si content and five with low Si. This is a separate approach compared to the AT analysis (Chapter 4) and aims to identify genes that are differentially expressed in discrete high and low leaf Si groups, compared to AT that is used to reveal SNP-trait associations and linear gene expression associations with the input trait. Additionally, the RNA and leaf samples were collected at the same time, compared to the AT analysis that used transcriptome data produced in 2016, approximately three years

prior to the collection of branch samples. This enables a more accurate genetic analysis of leaf Si.

The chapters aims to identify genes that are expressed at different levels in trees with high leaf Si compared to low Si. The genes detected from this analysis could be regulated by Si or contribute to leaf Si accumulation. At a genetics level, the analysis provides an improved understanding of the role of Si in tree species by revealing a selection of genes significantly associated with leaf Si content and provides a basis for future studies.

5.2 Methods

5.2.1 Sample selection and DNA extraction

Ash leaf samples were selected from amongst the Hawkhills (54.101° N, -1.178° W) *F. excelsior* collection (Chapter 3). Trees were rank ordered based on their silicon content from samples collected on the 12/07/2019. Five high Si accessions and five low Si accessions were selected for RNA extraction using frozen leaf tissue (-80ºC) previously collected on 12/07/2019. High and low Si groups had a mean Si content of 0.285% and 0.131% respectively, with a difference of 2.18x (Figure 5.1).

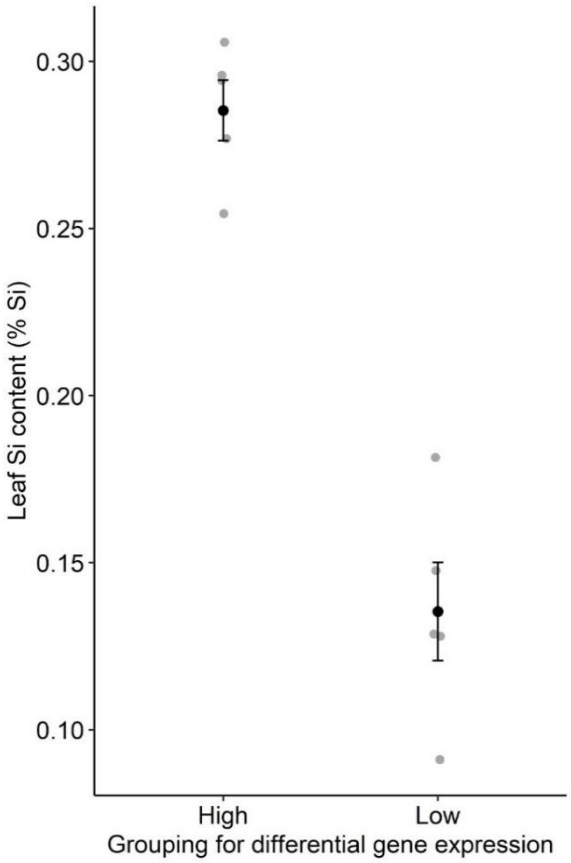

Figure 5.1. High and low leaf Si content grouping for differential gene expression analysis. Group means represented as black points and bars are +/- standard error. Grey points are individual values (n=5 per group).

Tissue homogenisation was conducted using a Qiagen tissuelyser machine set at a frequency of 30 beats s^{-1} for 30 seconds. The tissuelyser clamp and metal balls were pre-chilled overnight at -70ºC to prevent defrosting during homogenisation. Samples were immediately placed in dry ice and subsequently stored at -70ºC prior to RNA extraction.

RNA extraction followed the E.Z.N.A plant RNA kit protocol (version Sep. 2015), completing the additional DNase1 digestion steps. RNA was eluted into 50 µL DEPC water. RNA quality and concentration were then analysed using Nanodrop, retrieving 260:280 and 260:230 ratios, and plant RNA nano assay using Agilent 2100 bioanalyzer. Samples were diluted to the required concentrations prior to sequencing. Sequencing was carried out by Novogene Ltd using Illumina machines. Base calling, quality control, transcript counts and normalised quantification, as Fragments Per Kilobase of transcript per Million mapped reads (FPKM), were included in Novogene's service. The mapping of RNA reads to the reference genome

was achieved using HISAT2 v2.0.5 (Kim *et* al., 2019) and expression quantification using HTSeq v0.6.1 (Anders *et al*., 2015).

5.2.2 Differential gene expression analysis

The differential gene expression analysis was conducted using DESeq2 in RStudio 3.6.0 (Love *et al*., 2014). Transcript count data was pre-filtered by DESeq2 to remove 2442 genes from the analysis that had a sum transcript count value of less than or equal to one, across all ten leaf samples. The analysis tests for significant differences in individual gene expression between the high and low Si groups, outputting p values, $\log_2$ expression fold change (LFC) and LFC standard error for each gene. To produce more accurate estimates of expression fold change between the Si groups, DESeq2 was run using normal LFC shrinkage (Love *et al*., 2014) and bioinformatic analyses were conducted on the gene candidate using the same online tools and resources as Chapter 4.

5.3 Results

5.3.1 Identification of differentially expressed genes between high and low leaf Si groups

To detect genes that may control Si accumulation or be regulated by Si levels, ten ash trees were grouped in to high and low categories based on their leaf Si contents, calculated in Chapter 3. The high and low Si groups have a mean Si content of 0.285% and 0.131% respectively, with a fold difference of 2.18x (Fig. 5.1). After transcriptome sequencing, a differential gene expression analysis was then conducted using DESeq2. This analysis was carried out in addition to AT, as it aims to identify genetic differences between trees with high leaf Si accumulation and those with lower Si contents, instead of looking for linear associations between gene expression and Si content.

The analysis resulted in the identification of 17 genes that have significantly higher expression in the high leaf Si group (adjusted p value < 0.05 & log2 fold change > 1) and 18 genes expressed higher in the low Si group (adjusted p value < 0.05 & log2 fold change < -1). After processing the data using normal log fold change shrinkage, 12 (0.033%) genes have significantly higher expression in the high Si group and 10 (0.027%) in the low group (Table 5.1; Table 5.2).

The high expression, high Si genes are predicted to encode a diverse selection of proteins with functional roles in disease resistance, secondary metabolite biosynthesis, N-glycan processing, translation and regulation of gene expression (Table 5.1). The analysis also outputted multiple uncharacterised proteins, including FRAEX38873_v2_000083270, which encodes a protein with domain of unknown function (DUF538). This domain has been predicted to have chlorophyll hydrolysing activity and may be involved in plant stress responses (Gholizadeh, 2016).

One of the genes highly expressed in the high leaf Si group is predicted to be a disease resistance gene with moderate similarity to a probable disease resistance protein from *O. europaea* (BLASTx: 87% cover; 69% identity; XP_022897029.1) and an NB-ARC domain-containing protein when restricted to *A. thaliana* proteins. Analysis of conserved domains predicts the presence of an NB-ARC domain between amino acids 157 and 387. NB-ARC domains consist of three subdomains, nucleotide-binding

(NB), ARC1 and ARC2, and is thought to regulate expression of R genes via the NB domain (van Ooijen *et al*., 2008).

The expression level of two α-mannosidase genes, FRAEX38873_v2_000046140 and FRAEX38873_v2_000179130, are significantly associated with leaf Si content and are expressed at a greater level in the high Si group. The proteins encoded by these genes are predicted to be involved in the processing of N-glycans for protein glycosylation and display limited sequence similarity (EMBOSS Needle: 13.7% identity, 23.7% similarity, 45.8% gaps), possibly indicating roles in different stages of glycan processing. Despite these sequence differences, both genes have related sequences to an *O. europaea* α-mannosidase-like isoform X5 (XP_022885758.1) and to *A. thaliana g*lycosyl hydrolase family 38 proteins At3g26720, At5g66150 and At5g13980, which all encode class II lysosomal α-mannosidase enzymes rather than Golgi α-mannosidases (Strasser *et al*., 2006). This suggests a role in degradation of mannose-rich glycoproteins through their exoglycosidase activity (for review, Winchester, 2005).

Table 5.1. Differentially expressed genes between high and low silicon groups (Log2 fold change >1 and adjusted p value < 0.05). Annotations from BLASTx of the gene sequence. SE is the standard error of the log2-fold change.

Gene	Annotation	Adjusted p value	Log2 fold change	SE
FRAEX38873_v2_000019420	NB-ARC domain, disease resistance protein	0.037921	3.55	0.86
FRAEX38873_v2_000045230	LisH domain-containing protein	0.045164	2.49	0.61
FRAEX38873_v2_000046140	α-mannosidase	3.80E-07	1.39	0.21
FRAEX38873_v2_000083270	DUF538-containing protein	0.007018	1.49	0.32
FRAEX38873_v2_000174130	30S ribosomal protein 3-1, chloroplastic	0.006052	4.25	0.91
FRAEX38873_v2_000179130	α-mannosidase	0.027849	1.26	0.30
FRAEX38873_v2_000213290	Uncharacterised	0.045736	1.52	0.37
FRAEX38873_v2_000235420	β-caryophyllene synthase	0.001073	6.65	1.32
FRAEX38873_v2_000313840	LONGIFOLIA 1-like	0.031927	1.82	0.44
FRAEX38873_v2_000340110	Probable methyltransferase PMT11	0.000696	5.28	1.03
FRAEX38873_v2_000355230	Plastid terpene synthase	0.007018	3.71	0.81
FRAEX38873_v2_000371320	Golgin subfamily A member 4-like	0.031927	2.34	0.56

Two secondary metabolite biosynthesis genes are also expressed at a greater level in the high leaf Si group. These include the genes FRAEX38873_v2_000355230, with similarity to a plastid terpene synthase from *Osmanthus fragrans*, family Oleaceae (BLASTx: 90% cover, 88% identity; AMB57287.1), and FRAEX38873_v2_000235420, likely encoding an *O. europaea* β-caryophyllene synthase (BLASTx: 63% cover; 78% identity; XP_022858510.1). Protein translation tools predict that the CDS model of this gene contains two open reading frames, the first of which returns no BLASTx specific hits and the second returning an improved match to the β-caryophyllene synthase protein.

The predicted protein structure of the terpene synthase contains a terpene cyclase class 1 domain spanning amino acids 52 to 586. This is a member of the isoprenoid biosynthesis enzyme superfamily and may be involved in the synthesis of terpenoids, including monoterpenes, diterpenes or sesquiterpenes. The predicted amino acid sequence contains some similarity to *A. thaliana* terpene synthase-like 1,8-cineole At3g25830 (EMBOSS Needle: 34.7% identity; 53.8% similarity; 15.4% gaps), which is mainly involved in the synthesis of the monoterpene 1,8-cineole, consisting of 52% of reaction products, as well as several minor products including α-thujene (0.6%), α-pinene (1.9%), sabinene (14.5%), β-pinene (7.8%), myrcene (13.3%), limonene (4.0%), β-ocimene (2.7%), terpinolene (0.8%), and α-terpineol (2.4%) (Chen *et al.,* 2004).

The other metabolite biosynthesis gene, β-caryophyllene synthase, has the greatest log fold-change expression difference between high and low Si groups, with an LFC of 6.65 ± 1.32. Similar to the plastid terpene synthase candidate, the protein structure of the β-caryophyllene synthase gene is predicted to contain two class 1 terpene synthase domains. Additionally, it has sequence similarity to *A. thaliana* terpene synthase 21 (TPS21), which is involved in the production of the sesquiterpene β-caryophyllene, and minor products α-humulene and α-copaene, from the metabolite farnesyl pyrophosphate (FPP) (Chen *et al*., 2003). The classes of terpenes predicted to be produced by these two terpene synthase candidates are volatiles secreted from plants, for example by floral organs, and can be involved in herbivore defence, insect attraction and insect repellence.

Two of the genes displaying high expression in the high Si group are predicted to have roles in growth and development processes. These

include FRAEX38873_v2_000174130, a 30S ribosomal protein subunit expressed in chloroplasts (BLASTx: 99% cover; 92% identity; XP_022865301.1), and FRAEX38873_v2_000313840, a developmental protein LONGIFOLIA1 (BLASTx: 87% cover; 84% identity; XP_022883368.1), which contributes to leaf elongation in *Arabidopsis* (Lee *et al*., 2018).

The differential gene expression analysis resulted in identification of ten genes that have significantly lower expression in the high Si group and so could be downregulated by leaf Si or suppress Si accumulation. These genes have a variety of predicted functions, such as nitrogen transport, protein degradation, disease resistance and metabolite biosynthesis (Table 5.2).

FRAEX38873_v2_000115040 is predicted to encode a nitrogen transporter, NRT1/PTR family 5.10-like (BLASTx: 96% cover; 83% identity; XP_022850259.1). Conserved domain analysis revealed the presence of a major facilitator superfamily (MFS) domain, which is a large protein family involved in the transport of a wide range of molecules across cytoplasmic or intracellular membranes. The predicted protein sequence contains similarity to *A. thaliana* NPF5.10 (At1g22540), which induces nitrogen accumulation when expressed in *Xenopus* oocytes (Léran *et al*., 2015).

Two candidates are potentially involved in the degradation of proteins, including biogenesis of lysosome-related organelles complex 1 (BLOC1) subunit 1 (BLOS1), and F-box protein PP2-B10-like. FRAEX38873_v2_000398220, encoding the BLOS1 protein (BLASTx: 99% cover; 89% identity; XP_022847128.1), is a putative member of the eight subunit BLOC1 complex. The predicted ash BLOS1 protein contains a GCN5-like protein 1 domain (GCN5L1).

In *A. thaliana,* the BLOS1 protein is thought to interact with SORTING NEXIN (SNX) and BLOS2 proteins to enable endosomal transport from endosomes to lysosomes, suggesting roles in protein degradation via the vacuole degradative pathway. This has been hypothesised to be necessary for regulation of the cell surface receptors PIN-FORMED 1 (PIN1) and PIN2, which are key auxin efflux transporters contributing to plant growth and development (Cui *et al*., 2010).

Two disease related proteins are also identified with higher expression in the low Si group. One of these is predicted to encode a pathogenesis-related 5 (PR5)-like protein (BLASTx: 85% cover; 87% identity; XP_022860008.1), containing an antifungal thaumatin-like protein (TLP) domain between amino acid 37 and 253. This protein is similar to *A. thaliana* PR5-like protein ATLP-3, which has predicted antifungal activity and is induced by the defence phytohormone salicylic acid and bacterial pathogen *Pseudomonas syringae* (Hu and Reddy, 1997).

The other disease resistance gene, FRAEX38873_v2_000252560, has sequence similarity to *O. europaea* putative disease resistance protein RGA4 (BLASTx: 98% cover; 67% identity; XP_022849838). The predicted protein is highly leucine rich, with leucine-rich repeats (LRR) spanning between amino acid 10 and 172, corresponding to approximately 77% of the protein sequence. These LRR proteins, such as RGA4, are predicted to be R genes with functional roles in pathogen defence.

Finally, two metabolic enzymes displayed higher expression in the low leaf Si group. These include a flavanone 3-dioxygenase (FRAEX38873_v2_000325450; BLASTx: 99% cover, 69% identity; PIN00448.1), which is involved in flavonoid biosynthesis, and a thebaine 6-O-demethylase-like gene (FRAEX38873_v2_000364310; BLASTx: 74% cover; 88% identity; XP_022861165.1) that is likely involved in the alkaloid biosynthesis pathway.

The identification of 22 differentially expressed genes between high and low leaf Si groups provides an insight into the regulation and roles of Si in ash trees. Functional predictions of these genes, through bioinformatic analyses, suggests that expression of leaf developmental genes, disease resistance proteins and metabolite biosynthesis enzymes are associated with leaf Si content. However, further work is required to validate these genes and analyse their associations with Si in greater detail.

Table 5.2. Differentially expressed genes between high and low silicon groups (Log2 fold change < -1 and adjusted p value < 0.05). Annotations from BLASTx of the gene sequence. SE is the standard error of the log2-fold change.

Gene	Annotation	Adjusted p value	Log2 fold change	SE
FRAEX38873_v2_000031550	Uncharacterised	0.007628	-3.23	0.71
FRAEX38873_v2_000115040	NRT1/PTR family 5.10-like	0.046647	-3.44	0.85
FRAEX38873_v2_000149320	Triose phosphate/phosphate translocation, chloroplastic	0.007018	-1.19	0.26
FRAEX38873_v2_000234180	F-box protein PP2-B10-like	0.007018	-1.82	0.40
FRAEX38873_v2_000252560	Putative disease resistance protein RGA4	2.08E-07	-5.25	0.79
FRAEX38873_v2_000321620	Pathogenesis-related protein 5-like	0.019249	-1.71	0.39
FRAEX38873_v2_000322920	Uncharacterised	0.000532	-1.80	0.35
FRAEX38873_v2_000325450	Flavanone 3-dioxygenase	0.00013	-5.50	1.00
FRAEX38873_v2_000364310	Thebaine 6-O-demethylase-like	0.003735	-5.19	1.08
FRAEX38873_v2_000398220	Biogenesis of lysosome-related organelles complex 1 subunit 1	0.036673	-1.43	0.35

5.4 Discussion

The aim of this chapter was to identify genes that are expressed at different levels in trees with relatively high leaf Si contents compared to those with low Si. This detects genes that could be regulated by Si or contribute to leaf Si accumulation and is useful for providing novel understanding of the role of Si in tree species at a genetics level. The differential gene expression analysis between high and low leaf Si groups revealed 22 genes with significantly different expression levels. These were predicted to encode a diverse range of proteins, with functional roles involved in disease resistance, secondary metabolite biosynthesis, regulation of gene expression and other cellular processes.

Multiple defence-associated genes were identified as displaying significant differential expression between high and low leaf Si groups. A positive association was detected between leaf Si and the expression of an NB-ARC domain, disease resistance protein. This domain is predicted to regulate the activity of pathogen resistance genes through nucleotide binding (van Ooijen *et al*., 2008). Additionally, two defence-related proteins in ash, PR5-like and RGA4, displayed a negative association with leaf Si content. The PR5-like protein contained sequence similarity to *A. thaliana ATLP-3*, also encoding a PR5-like protein. This gene is known to be induced by the bacterial pathogen *Pseudomonas syringae*, as well as salicylic acid (Hu and Reddy, 1997), often associated with systemic acquired resistance (SAR). The same study also predicted antifungal activity for *ATLP-3*, as this protein was able to inhibit *Candida albicans* growth *in vitro*.

The other disease resistance protein RGA4 contains a high proportion of LRRs and is likely an R protein that recognises pathogen avirulence proteins (AVR) to initiate a defence response. Specifically, in rice, RGA4 is predicted to respond to the AVR genes AVR1-CO39 and AVR-Pia from the rice blast fungal pathogen *M. oryzae*, which are initially recognised by R gene RGA5 (César *et al*., 2013; 2014). Through loss of function mutations and complementation experiments, RGA4 and RGA5 have been demonstrated to be involved in AVR1-CO39–triggered resistance to *M. oryzae* (César *et al*., 2013).

The differential expression of several pathogen resistance genes suggests that plant Si content could influence the expression profile of pathogen

defence-associated genes and may be a factor contributing to the beneficial effects of Si on plant disease resistance. Although further study investigating these associations would be required to draw conclusions. Similar relationships have been observed between Si and defence-associated genes in rice and tomato plants when exposed to biotic stress (reviewed by Manivannan and Ahn, 2017). Specifically, when tomato plants were exposed to the bacterium *Ralstonia solanacearum*, defence-associated genes such as disease resistance response protein, WRKY1 and ferritin-related protein, were significantly enriched in Si supplemented plants compared to unfertilised controls. The Si fertilised plants also displayed reduced disease severity (Ghareeb *et al*., 2011), illustrating the link between disease resistance and Si.

Alternatively, it is possible that defence-associated genes influence Si uptake and deposition. Previous links have been demonstrated between jasmonic acid signalling and the expression of silicon transporters in rice (Ye *et al*., 2013). Rice plants treated with methyl jasmonate (MeJA) displayed an increase in leaf Si content by 65.5%, whereas knockdown of *CORONATINE INSENSITIVE1*, a key jasmonic acid signalling gene, resulted in reduced Si content for MeJA-treated and control plants (Ye *et al*., 2013). This suggests a potential role of defence-hormone signalling in Si deposition in rice.

Several metabolite biosynthesis genes were found to be differentially expressed, with potential roles in plant defence against pathogens and herbivores. One of these was predicted to encode a β-caryophyllene synthase. This protein is involved in the production of the sesquiterpene β-caryophyllene from FPP, which is a volatile secondary metabolite released at the roots and leaves in response to herbivore stimulation. For example, in lima bean (*Phaseolus lunatus*) a β-caryophyllene synthase was found to be upregulated after treatment with spider mites and alamethicin, a volatile elicitor from the fungus *Trichoderma viride* (Li *et al*., 2017). Additionally, high repellent activity has been demonstrated for β-caryophyllene against spider mites (Camara *et al*., 2015).

However, a study investigating the effect of Si supplementation on volatile metabolite production in rice found no association between Si and β-caryophyllene levels (Liu *et al*., 2017). This appears to contradict the results from this investigation which discovered higher β-caryophyllene synthase

expression in the presence of high leaf Si levels. Therefore, further study would be required to confirm this association and could be achieved by measuring β-caryophyllene synthase gene expression and β-caryophyllene content in ash trees under control conditions with and without Si supplementation.

The other secondary metabolite synthase gene expressed at a greater level in the high leaf Si group was predicted to encode a terpene synthase, involved in the production of the monoterpene 1,8-cineole. This metabolite is found in high abundance in the essential oils of eucalyptus species where it is hypothesised to contribute antimicrobial and antifungal activity (Marzoug *et al*., 2011).

The association between leaf Si and the expression of secondary metabolite biosynthesis enzymes provides evidence towards the role of Si in plant metabolism. As these volatile metabolites are associated with pathogen defence and insect repellence, it offers an additional explanation for the beneficial effects of Si in promoting disease resistance.

Similar to the results from the AT analysis of branch Si content (Chapter 4), several of the differentially expressed genes are predicted to contribute towards root and leaf development. Two of these genes were predicted to encode class II lysosomal α-mannosidases involved in degradation of mannose-rich glycoproteins in lysosomes. N-linked glycosylation is the attachment of branched oligosaccharide structures to proteins and can involve the addition of mannose, N-Acetylglucosamine, xylose, fucose and galactose monomers. The processing of mannose in N-glycan structures is important in developmental processes. For example, in *Arabidopsis,* three Golgi α-mannosidases (*mns1, 2* and *3*) appear to contribute to correct root formation as loss of glycan processing from $Man_9GlcNAc_2$ to smaller structures, via knockout of these genes, results in mutants displaying short root phenotypes with increased lateral branching (Liebminger *et al*., 2009). Additionally, N-glycan processing may play a role in cell wall development, as mutants display lower levels of highly methylated homogalacturonan compared to wild-type plants.

One of the genes expressed at a higher level in the high leaf Si group was predicted to encode a plastid-specific 30S ribosomal protein 3-1 (PSRP3). In *Arabidopsis, PSRP*3 expression is known to affect leaf growth, leaf cell differentiation, chlorophyll content and translation of the key photosynthetic

components (Tiller *et al*., 2012). Knockout mutations of this gene have clear reductions in plant size and yellowing of leaves when compared to wild-type plants, indicating reduced chlorophyll content. Although, the PSRP3 gene appears to have some functional redundancy or has a less essential role, based on more severe mutant phenotypes from other PSRP genes.

The *A. thaliana psrp3* mutant has pronounced effects on leaf tissue anatomy, with a shortening of palisade cells and smaller mesophyll cells being observed. The mutant also displays reduced levels of chlorophyll and reduced expression of key photosynthetic components, including photosystem I, II (PSI and II) and cytochrome b 6f, indicating the critical role this gene has in photosynthesis and leaf development. Additionally, previous differential gene expression analyses have revealed that components of PSI and PSII, including PsbY PSII subunit and PsaH PSI subunit, are upregulated in Si supplemented rice plants under zinc toxicity conditions compared to Si-deprived controls (Song *et al*., 2014). The authors hypothesised that this could improve the efficiency of photosynthesis and counter the negative impacts of abiotic and biotic stresses. Therefore, in ash trees, this gene could be involved in leaf Si deposition through increased transpiration via photosynthesis or by enabling correct leaf development. Alternatively, the regulation of *PSRP3* by Si may result in improve stress responses through enhanced photosynthesis efficiency.

Another differentially expressed gene associated with leaf development is the LONGIFOLIA1 protein, which displays greater expression in the high leaf Si group. Previous studies have revealed that this protein primarily functions to promote leaf cell elongation during development, although there is functional redundancy between four *A. thaliana* LNG genes (LNG1-4) (Lee *et al*., 2018). Specifically, these genes are involved in leaf elongation along the proximodistal axis, with knockout mutants *lng* quadruple (*lnq*), *lng1/2/3* and *lng1/2/4* triple displaying reduced leaf length compared to wild type, with leaf width remaining similar (Lee *et al*., 2018). The *lnq* quadruple knockdown has the most severe phenotype, with a reduction in leaf length of 6.9 mm (36%). A closer look at mutant leaf structure showed that palisade cells were approximately 30% shorter in the proximodistal direction, explaining the reduction in leaf length. Conversely, an *LNG1* overexpressing line displayed long leaves and petioles, caused by a 51% increase in epidermal cell length (Lee *et al*., 2006).

The molecular and genetic interactions of this protein have not been fully investigated. However, in *Arabidopsis* seed coat epidermal cells, this protein is involved in microtubule organisation and interacts with CELLULOSE SYNTHASE 3 to ensure correct cellulose orientation, demonstrated through analysis of mutant lines (Yang *et al*., 2018). Additionally, based on expression changes between *lnq* knockout, *lng1* overexpression and wild-type plants, LNG1 is hypothesised to positively regulate the expression of xyloglucan endotransglucosylase (XEH) /hydrolase (XTH) genes (Lee *et al*., 2018). These protein families are important in the deposition of xyloglucans in the plant cell walls, as *A. thaliana xth31* mutants contain significantly reduced levels of xyloglucan oligosaccharides in cell walls digested with xyloglucanase (Zhu *et al*., 2012).

The influence this gene has on leaf morphology may affect transpiration and therefore Si accumulation. Additionally, the association between LNG genes and cell wall development, specifically deposition of cellulose and xyloglucan, could affect Si accumulation at the cell wall.

Finally, the ash *BLOS1* gene was negatively associated with leaf Si content, with lower expression in the high leaf Si group. The protein encoded by this gene is known to interact with SNX and BLOS2 proteins in *Arabidopsis,* to enable endosomal transport from endosomes to lysosomes (Cui *et al*., 2010). This SNX-BLOS1 interaction occurs via a GCN5L domain, which is also present in the *F. excelsior* homologue, suggesting that a similar interaction could occur in ash.

It is hypothesised that the protein complex is involved in the degradation of PIN1 and PIN2, two key auxin efflux transporters. The protein quantity of PIN1 and PIN2 has been demonstrated to be greater in *A. thaliana* mutants with RNAi knockdown of *BLOS1*, compared to wild-type plants, without an increase in PIN1/2 transcripts (Cui *et al*., 2010). Therefore, linking *BLOS1* activity to regulation of auxin signalling, which could influence the growth and development of plants. Some evidence of these effects was highlighted in the *blos1* RNAi mutant, which displayed significantly longer primary roots and increased lateral root branching. The promotion of root growth at low BLOS1 expression levels may influence Si uptake from the environment and could indicate why higher leaf Si is detected in ash trees with reduced *BLOS1* transcripts.

From these results, it appears that leaf Si uptake is influenced by root and leaf development, via genes such as *PSRP3*, *LNG1* and *BLOC1*. Si deposition may be affected by α-mannosidases and LNG1 proteins, which have predicted roles in cell wall development and composition, factors which could affect Si associated with the cell wall. These results are similar to the candidates identified from the branch Si AT experiment (Chapter 4), which identified several genes that associated with branch Si content and have predicted functions in promoting leaf, root and vascular tissue development.

In addition to the developmental genes, several disease related genes displayed significant differential expression between the leaf Si groups. Therefore, Si could alter expression of disease-resistance genes or the biosynthesis of secondary metabolites with anti-herbivore properties. This could explain the beneficial effect of Si in promoting disease and pest resistance.

The identification of 22 significantly differentially expressed genes associated with leaf Si content provides a foundation for future studies validating and further investigating the biological significance of these candidates. Based on the available literature, there appear to be no direct, proven links between Si and the candidates identified from this analysis. However, the expression of several defence-associated genes and photosynthetic components from previous analyses have been associated with Si supplementation under stress conditions (Ghareeb *et al*., 2011; Song *et al*., 2014). Given that these are similar to some of the candidates identified from the analysis in this chapter, it would be reasonable to investigate these candidates in greater detail. To further analyse the differentially expressed genes, it would be useful to analyse ash trees under controlled conditions with and without silicon supplementation, followed by quantification of the candidate gene expression. This will allow the association between gene and Si to be investigated in more detail. If the expression level of the candidates changes after Si fertilisation then these genes are likely regulated by Si in ash. Otherwise, they may be false positive results or influence Si accumulation.

Chapter 6 – Project discussion and conclusions

The aims of this project were to improve our knowledge of Si in ash trees, identify gene markers for Si content, begin to understand the genetics associated with Si in ash, and to improve the genetic resources available. These aims were achieved through XRF Si measurements of a diverse panel of ash trees, genetic analyses including differential gene expression analysis and associative transcriptomics, and through the assembly of long-read sequence data.

The previous Sollars *et al*. (2017) *F. excelsior* genome has been used to investigate the population and genetic structures of ash, and to predict susceptibility to ash dieback using gene markers. However, this assembly remains highly fragmented, containing approximately 89,000 relatively short contig sequences, which restricts the effectiveness of the assembly for use in genome-wide genetic analyses. Therefore, using long-read ONT sequencing the aim was to produce a highly contiguous genome with fewer and longer contig sequences. After assembly, quality control and filtering processes, a more contiguous genome was generated, which was composed of a relatively low number of sequences compared to the previous assembly. Comparisons between the two assemblies revealed approximately 86,000 fewer sequences for the ONT genome and a 5.5x greater N_{50} value, demonstrating substantially improved genome contiguity.

The improved assembly was used to reorder the existing ash transcriptome, which could then be implemented into the AT analysis to produce Manhattan plots showing the degree of marker association to the trait of interest in each contig of the genome. Although no significant SNP markers were detected for the traits analysed in this project, the mapping of the transcriptome to a more contiguous genome should improve the analysis of SNPs, as associated markers are not necessarily found in the genes controlling traits of interest, but can instead be located nearby in the genome.

However, analysis of the assembly revealed that it requires further work to reach the quality of recent ONT draft assemblies. For example, the *L. japonica* genome was produced to a much higher contiguity with an N_{50} of 84 Mbp (Pu *et al*., 2020), compared to the 568 Kbp N_{50} for the *F. excelsior* genome. In addition to long-read sequences, the *L. japonica* genome used short reads, produced using Illumina, and Hi-C sequencing to produce

pseudochromosome predictions. These methods could be applied to the *F. excelsior* genome to further improve the assembly contiguity and accuracy.

To improve our understanding of Si in ash trees various Si measurements were conducted. The degree of Si uptake and deposition have been investigated for a large variety of species, across angiosperms and gymnosperms (Hodson *et al*., 2005). However, the extent of Si uptake in tree species has been somewhat overlooked compared to the main study species in the field, including rice and horsetail. Studies such as Cornelis *et al*. (2010) have provided an insight into the Si content of some deciduous and coniferous trees, including European beech, oak, black pine, Douglas fir and Norway spruce, and reveals the distribution of Si within the tree leaves, branches, stem wood and stem bark for these species. However, the results of these studies demonstrate the large amount of variation that occurs between plant families and at a species level. Therefore, it was important to first conduct Si measurements in ash to determine the extent of Si uptake, concentration of deposited Si and the distribution between tissues.

The analysis of XRF Si data generated from leaf and branch samples revealed the presence of low-level Si accumulation, which favoured deposition in leaf tissue compared to branches, and demonstrated the accumulation of Si as the leaves matured. These results could suggest a passive mechanism of Si uptake, driven by the transpiration stream, as proposed for other tree species including *Citrus sinensis* and *C. reticulata* (Mvondo-She and Marais, 2019). This hypothesis would also explain the reduction in Si deposition for trees severely infected with ash dieback, as crown dieback and infection of vascular tissue would result in disrupted transpiration.

Although these results advance our knowledge of Si in ash trees, microscopy techniques such as scanning electron microscopy, light microscopy with histological stains, or Raman analysis (Blecher *et al*., 2012) could be used to visualise the structure and deposition pattern of Si in various ash tissues. This could provide additional information about how the deposition process occurs in ash and may further advance our knowledge of the role of Si in tree species.

Following identification of Si uptake and deposition in ash, the project aimed to investigate the genetics associated with Si content. Previous

molecular and genetics studies focusing on Si in plants have identified the presence of Si transporters in monocots such as rice (Ma *et al*., 2006; 2007) and dicots including soybean (Deshmukh *et al*., 2013). Molecules potentially linked to Si deposition, including callose (Law and Exley, 2011), pectin and hemicellulose (Gierlinger *et al*., 2008), are also identified in a limited number of species. The influence of Si on transcription profiles has also been analysed, with Si hypothesised to associate with pathogenesis-related genes, hormone biosynthesis and stress-related genes (Jiang *et al*., 2019). However, the genetics associated with Si in tree species has not been investigated. Therefore, this project used genomic methods including associative transcriptomics of branch Si data and differential gene expression analysis of leaf Si data to identify genetic markers associated with Si content, improving our understanding of the genes linked to Si in ash trees.

The associative transcriptomics analysis of branch Si data revealed nine significantly associated GEMs. The majority of these are predicted to influence the growth and development of leaf, root and vascular tissue, and may be involved in processes such as the control of cell wall composition, cell elongation and microtubule formation. This could be consistent with a passive mechanism of Si accumulation as trees with developed root, leaf and vascular tissue may be able to passively uptake more Si.

The differential expression results revealed 22 significantly differentially expressed genes associated with leaf Si content. A selection of these genes have predicted roles in developmental processes, are pathogen defence-associated genes or are involved in the biosynthesis of secondary metabolites with anti-herbivore properties. This could explain the beneficial effect of Si in promoting disease resistance through regulation of defence-associated genes. Alternatively, Si accumulation may be influenced by these gene families, revealing potential active components of Si uptake and deposition in ash. Both of these scenarios have been demonstrated in rice where plants primed with Si increased expression of defence-associated genes upon insect attack, compared to plants without Si supplementation, and the accumulation of Si was promoted by jasmonic acid signalling, with mutants defective in jasmonic acid biosynthesis and recognition displayed reduced Si content compared to wild-type plants (Ye *et al*., 2013). However, further work will be needed to validate these candidates to confirm they are not false positive associations and could be achieved by

checking the associations in ash trees that are independent from the initial analysis. This could then be followed by an investigation into the relationship between Si and the candidate genes to determine whether Si accumulation is influenced by the expression of defence-associated genes or are instead regulated by Si in the tree.

The results of the two analyses suggest that genetics could be involved in Si uptake and deposition, or that gene expression is regulated by Si. Leaf and root development were identified as processes potentially linked to Si content, with the identification of genes such as the *LNG1*-like gene involved in leaf development (Lee *et al*., 2018) and *IREH1*-like gene controlling root elongation (Oyama *et al*., 2002). These processes could be associated with Si accumulation by influencing silicic acid uptake from the soil or transport and condensation by increased transpiration.

Additionally, the analyses identified genes that could influence cell wall composition, including the kinase PERK9 that may promote callose deposition, as demonstrated for the related protein PERK1 (Haffani *et al*., 2006), and LNG1 that is hypothesised to positively regulate the expression of XEH and XTH genes (Lee *et al*., 2018), which influence the xyloglucan content of the cell wall (Zhu *et al*., 2012), and also controls the deposition of cellulose in leaves (Yang *et al*., 2018). Cell-wall associated Si is visualised in other plant species (Blecher et al., 2012) and polysaccharides including cellulose (Perry and Lu, 1992), callose (Law and Exley, 2011), pectin and hemicellulose (Gierlinger *et al*., 2008) are hypothesised to be involved in Si deposition and organisation. Therefore, the results suggest that genes regulating the composition of the cell wall could be important in promoting deposition of Si in the leaf and branch tissue.

Overall, the results of this project provide significant advances in our knowledge of Si physiology in ash and the potential underlying genetics associated with biological silica in tree species. The outcome of the genome sequencing project provides a useful resource, beneficial for future genetics studies in ash trees, and could be important in solving the problems caused by ash dieback and emerald ash borer. Additionally, the results of the AT and differential gene expression analyses offer an understanding of the genetics associated with Si in ash trees, highlights some of the biological processes influencing or affected by Si and provides

a basis for future work focusing on validating and functionally analysing the candidate genes.

Word counts: Abstract (288); Thesis (24,007)

Abbreviations

Ash dieback (ADB)

Associative transcriptomics (AT)

Avirulence proteins (AVR)

Coding sequence (CDS)

domain of unknown function (DUF)

Effector-triggered immunity (ETI)

Emerald ash borer (EAB)

False discovery rate (FDR)

Farnesyl pyrophosphate (FPP)

Fragments Per Kilobase of transcript per Million mapped reads (FPKM)

Gene expression marker (GEM)

Genome-wide association studies (GWAS)

Leucine-rich repeat (LRR)

Log2 expression fold change (LFC)

Median read depth (MRD)

Microbe-associated molecular pattern (MAMP)

Nucleotide binding (NB)

Oxford Nanopore Technologies (ONT)

Pathogen-associated molecular pattern (PAMP)

Pathogenesis-related gene (PR)

Pattern recognition receptor (PRR)

Pattern-triggered immunity (PTI)

Photosystem (PS)

Reactive oxygen species (ROS)

Reads Per Kilobase of transcript per Million mapped reads (RPKM)

Selectivity filter (SF)

Single nucleotide polymorphism (SNP)

Systemic acquired resistance (SAR)

X-ray fluorescence (XRF)

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25: 3389-3402.

Anders, S., Pyl, P. T. and Huber, W. (2015). HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166-169. doi: 10.1093/bioinformatics/btu638

Bari, R. and Jones, J. D. G. (2009). Role of plant hormones in plant defence responses. *Plant Mol. Biol.* 69: 473-488. doi: 10.1007/s11103-008-9435-0

Bartoli, F. (1983) The biological cycle of silicon in two temperate forest ecosystems. *Ecol. Bull.* 35: 469-476.

Belton, J., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y. and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* 58: 268-276. doi: 10.1016/j.ymeth.2012.05.001

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* 57: 289-300.

Bernhardt, A., Lechner, E., Hano, P., Schade, V. Dieterle, M., Ander, M. *et al*. (2006). CUL4 associates with DDB1 and DET1 and its downregulation affects diverse aspects of development in *Arabidopsis thaliana. Plant J.* 47: 591-603. doi: 0.1111/j.1365-313X.2006.02810.x

Bhuiyan, N. H., Selvaraj, G., Wei, Y. and King J. (2009). Gene expression profiling and silencing reveal that monolignol biosynthesis plays a critical role in penetration defence in wheat against powdery mildew invasion. *J. Exp. Bot.* 60: 509-521. doi: 10.1093/jxb/ern290

Bigeard, J., Colcombet, J. and Hirt, H. (2015). Signaling mechanisms in pattern-triggered immunity (PTI). *Mol. Plant* 8: 521-539. doi: 10.1016/j.molp.2014.12.022

Biological Records Centre (2008). Database of insects and their food plants. brc.ac.uk/DBIF

Blecher, I. C., Seidel, R., Thomann, R. and Speck, T. (2012). Comparison of different methods for the detection of silica inclusions in plant tissues. *Int. J. Plant Sci.* 173: 229-238. doi: 10.1086/663969

Bögre, L., Ökrész, L., Henriques, R. and Anthony, R. G. (2003). Growth signalling pathways in Arabidopsis and the AGC protein kinases. *Trends Plant Sci.* 9: 424-431. doi: 10.1016/S1360-1385(03)00188-2

Brugiére, T. and Exley, C. (2017). Callose-associated silica deposition in Arabidopsis. *J. Trace Elem. Med. Bio.* 39: 86-90. doi: 10.1016/j.jtemb.2016.08.005

Camara, C. A. G., Akhtar, Y., Isman, M. B., Seffrin, R. C. and Born, F. S. (2015). Repellent activity of essential oils from two species of Citrus against *Tetranychus urticae* in the laboratory and greenhouse. *Crop Prot.* 74: 110-115. doi: 10.1016/j.cropro.2015.04.014

Césari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A. *et al*. (2013). The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell* 25: 1463-1481. doi: 10.1105/tpc.112.107201

Césari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y. *et al*. (2014). The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *EMBO J.* 33: 1941-1959. doi: 10.15252/embj.201487923

Chen, F., Tholl, D., D'Auriam J. C., Farooq, A., Pichersky, E. and Gershenzon, J. (2003). biosynthesis and emission of terpenoid volatiles from Arabidopsis flowers. *Plant Cell* 15: 481-494. doi: 10.1105/tpc.007989

Chen, F., Ro, D., Petri, J., Gershenzon, J., Bohlmann, J., Pichersky, E. *et al*. (2004). Characterization of a root-specific *Arabidopsis* terpene synthase responsible for the formation of the volatile monoterpene 1,8-Cineole. *Plant Physiol.* 135: 1956-1966. doi: 10.1104/pp.104.044388

Chérif, M., Menzies, J. G., Benhamou, N. and Bélanger, R. R. (1992). Studies of silicon distribution in wounded and *Pythium ultimum* infected cucumber plants. *Physiol. Mol. Plant P.* 41: 371-385. doi: 10.1016/0885-5765(92)90022-N

Chiba, Y., Mitani, N., Yamaji, N. and Ma, J. F. (2009). HvLsi1 is a silicon influx transporter in barley. *Plant J.* 57: 810-818. doi: 10.1111/j.1365-313X.2008.03728.x

Cleary, M. R., Daniel, G. and Stenlid, J. (2013). Light and scanning electron microscopy studies of the early infection stages of *Hymenoscyphus pseudoalbidus* on *Fraxinus excelsior*. *Plant Pathol.* 62: 1294-1301. doi: 10.1111/ppa.12048

Coker, T. L. R., Rozsypálek, J., Edwards, A., Harwood, T. P., Butfoy, L. & Buggs, R. J. A. (2018). Estimating mortality rates of European ash (*Fraxinus excelsior*) under the ash dieback (*Hymenoscyphus fraxineus*) epidemic. *Plants People Planet* 1:48-58. doi: 10.1002/ppp3.11

Conrath, U. (2006). Systemic acquired resistance. *Plant Signal. Behav.* 1: 179-184. doi: 10.4161/psb.1.4.3221

Cornelis, J. T., Ranger, J., Iserentant, A. & Delvaux, B. (2010). Tree species impact the terrestrial cycle of silicon through various uptakes. *Biogeochemistry* 97: 231-245. doi: 10.1007/s10533-009-9369-x

Cui, Y., Li, X., Chen, Q., He, X., Yang, Q., Zhang, A. *et al*. (2010). BLOS1, a putative BLOC-1 subunit, interacts with SNX1 and modulates root growth in *Arabidopsis. J. Cell Sci.* 123: 3727-3733. doi: 10.1242/jcs.069732

Currie, H. A. and Perry, C. C. (2007). Silica in plants: Biological, biochemical and chemical studies. *Ann. Bot.* 100: 1383-1389. doi: 10.1093/aob/mcm247

Daniell, H., Lin, C. S., Yu, M. and Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17: 134. doi: 10.1186/s13059-016-1004-2

Deshmukh, R. K., Vivancos, J., Guérin, V., Sonah, H., Labbé, C., Belzile, F. *et al*. (2013). Identification and functional characterization of silicon transporters in soybean using comparative genomics of major intrinsic proteins in *Arabidopsis* and rice. *Plant Mol. Biol.* 83: 303-315. doi: 10.1007/s11103-013-0087-3

Deshmukh, R. K., Vivancos, J., Ramakrishnan, G., Guérin, V., Carpentier, G., Sonah, H. *et al*. (2015). A precise spacing between the NPA domains of aquaporins is essential for silicon permeability in plants. *Plant J.* 83: 489-500. doi: 10.1111/tpj.12904

Dobrowolska, D., Hein, S., Oosterbaan, A., Wagner, S., Clark, J. and Skovsgaard, J. P. (2011). A review of European ash (Fraxinus excelsior L.): implications for silviculture. *Int. J. For. Res.* 84: 133-148. doi: 10.1093/forestry/cpr001

Durak, G. M., Taylor, A. R., Walker, C. E., Probert, I., de Vargas, C., Audic, S. *et al*. (2016). A role for diatom-like silicon transporters in calcifying coccolithophores. *Nat. Commun.* 7: 10543. doi: 10.1038/ncomms10543

Epstein, E. (1994). The anomaly of silicon in plant biology. *PNAS* 91: 11-17. doi: 10.1073/pnas.91.1.11

Fang, C., Zhang, P., Jian, X., Chen, W., Lin, H., Li, Y. *et al*. (2017). Overexpression of Lsi1 in cold-sensitive rice mediates transcriptional regulatory networks and enhances resistance to chilling stress. *Plant Science* 262: 115-126. doi: j.plantsci.2017.06.002

Forestry Commission (2013). NFI Preliminary Report: NFI preliminary estimates of quantities of broadleaved species in British woodlands, with special focus on ash. National Forest Inventory, (Forestry Commission, Edinburgh).

Fortunato, A. A., da Silva, W. L. and Rodrigues, F. A. (2014). Phenylpropanoid pathway is potentiated by silicon in the roots of banana plants during the infection process of *Fusarium oxysporum* f. sp. cubense. *Phytopathology* 104: 597-603. doi: 10.1094/PHYTO-07-13-0203-R.

Fraysse, F., Pokrovsky, O. S., Schott, J. and Meunier, J. D. (2006). Surface properties, solubility and dissolution kinetics of bamboo phytoliths. *Geochim. Cosmochim. Ac.* 70:1939–1951. doi: 10.1016/j.gca.2005.12.025

Fraysse, F., Pokrovsky, O. S., Schott, J. and Meunier, J. D. (2009). Surface chemistry and reactivity of plant phytoliths in aqueous solutions. *Chemical Geology* 258: 197-206. doi: 10.1016/j.chemgeo.2008.10.003

Ghareeb, H., Bozsó, Z., Ott, P. G., Repenning, C., Stahl, F. and Wydra, K. (2011). Transcriptome of silicon-induced resistance against *Ralstonia solanacearum* in the silicon non-accumulator tomato implicates priming effect. *Physiol. Mol. Plant P.* 75: 83-89. doi: 10.1016/j.pmpp.2010.11.004

Gholizadeh, A. (2016). DUF538 protein superfamily is predicted to be chlorophyll hydrolyzing enzymes in plants. *Physiol. Mol. Biol. Plants* 22: 77-85. doi: 10.1007/s12298-015-0331-1

Gierlinger, N., Sapei, L. and Paris, O. (2008). Insights into the chemical composition of *Equisetum hyemale* by high resolution Raman imaging. *Planta* 227: 969-980. doi: 10.1007/s00425-007-0671-3

Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. C. and McCombie, W. R. (2015). Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genomes Res.* 25: 1750-1756. doi: 10.1101/gr.191395.115

Grašič, M., Škoda, B., Golob, A., Vogel-Mikuš, K. and Gaberščik, A. (2019). Barley and spelt differ in leaf silicon content and other leaf traits. *Biologia* 74: 929–939. doi: https://doi.org/10.2478/s11756-019-00227-w

Gross, A., Holdenrieder, O., Pautasso, M., Queloz, V. and Sieber, T. N. (2014). *Hymenoscyphus pseudoalbidus*, the causal agent of European ash dieback. *Mol. Plant Pathol.* 15: 5-21. doi: 10.1111/mpp.12073

Guerriero, G., Hausman, J. and Legay, S. (2016). Silicon and the plant extracellular matrix. *Front. Plant Sci.* 7: 463. doi: 10.3389/fpls.2016.00463

Guerriero, G., Law, C., Stokes, I., Moore, K. L. and Exley, C. (2018). Rough and tough. How does silicic acid protect horsetail from fungal infection? *J. Trace Elem. Med. Bio.* 47: 45-52. doi: 10.1016/j.jtemb.2018.01.015

Haffani, Y., Silva-Gagliardi, N., Sewter, S., Aldea, M. G., Zhao, Z., Nakhamchik, A. *et al*. (2006). Altered expression of PERK receptor kinases in *Arabidopsis* leads to changes in growth and floral organ formation. *Plant Signal. Behav.* 1: 252-260. doi: 10.4161/psb.1.5.3324

Han, Y., Wen, J., Peng, Z., Zhang, D. and Hou, M. (2018). Effects of silicon amendment on the occurrence of rice insect pests and diseases in a field test. *J. Integr. Agric*. 17: 2172-2181. doi: 10.1016/S2095-3119(18)62035-0

Harrison, C. C. (1996). Evidence for intramineral macromolecules containing protein from plant silicas. *Phytochemistry* 41: 37-42. doi: 10.1016/0031-9422(95)00576-5

Havlickova, L., He, Z., Wang, L., Langer, S., Harper, A. L., Kaur, H. *et al*. (2017). Validation of an updated associative transcriptomics platform for the polyploid crop species *Brassica napus* by dissection of the genetic architecture of erucic acid and tocopherol isoform variation in seeds. *Plant J.* 93: 181-192. doi: 10.1111/tpj.13767

Herms, D. A. and McCullough, D. G. (2014). Emerald ash borer invasion of north America: history, biology, ecology, impacts, and management. *Annu. Rev. Entomol.* 59: 13-30. doi: 10.1146/annurev-ento-011613-162051

Hildebrand, M., Dahlin, K. and Volcani, B. E. (1998). Characterization of a silicon transporter gene family in *Cylindrotheca fusiformis*: sequences, expression analysis, and identification of homologs in other diatoms. *Mol. Gen. Genet.* 260: 480-486. doi: 10.1007/s004380050920

Hill, L., Jones, G., Atkinson, N., Hector, A., Hemery, G. and Brown, N. (2019). The £15 billion cost of ash dieback in Britain. *Curr. Biol.* 29: R315-R316. doi: 10.1016/j.cub.2019.03.033

Hodson, M. J., White, P. J., Mead, A. and Broadley, M. R. (2005). Phylogenetic variation in the silicon composition of plants. *Ann. Bot.* 96: 1027-1046. doi: 10.1093/aob/mci255

Hu, X. and Reddy, A. S. N. (1997). Cloning and expression of a PR5-like protein from *Arabidopsis*: inhibition of fungal growth by bacterially expressed protein. *Plant Mol. Biol.* 34: 949-959. doi: 10.1023/A:1005893119263

Humphrey, T. V., Haasen, K. E., Aldea-Brydges, M. G., Sun, H., Zayed, Y., Indriolo, E. *et al*. (2015). PERK–KIPK–KCBP signalling negatively regulates root growth in *Arabidopsis thaliana. Journal of Experimental Botany* 66: 71-83. doi: 10.1093/jxb/eru390

Inoue, T., Okane, I., Ishiga, Y., Degawa, Y., Hosoya, T. and Yamaoka, Y. (2019). The life cycle of *Hymenoscyphus fraxineus* on Manchurian ash, *Fraxinus mandshurica*, in Japan. *Mycoscience* 60: 89-94. doi: 10.1016/j.myc.2018.12.003

Jiang, N., Fan, X., Lin, W., Wang, G. and Cai, K. (2019). Transcriptome analysis reveals new insights into the bacterial wilt resistance mechanism mediated by silicon in tomato. *Int. J. Mol. Sci.* 20: 761. doi: 10.3390/ijms20030761

Johnson, L. K., Sahasrabudhe, R., Gill, J. A., Roach, J. L., Froenicke, L., Brown, C. T. *et al*. (2020). Draft genome assemblies using sequencing reads from Oxford Nanopore Technology and Illumina platforms for four species of North American *Fundulus* killifish. *GigaScience* 9: giaa067. doi: 10.1093/gigascience/giaa067

Kablan, L., Lagauche, A., Delvaux, B. and Legreve, A. (2012). Silicon reduces black sigatoka development in banana. *Plant Dis*. 96: 273-278. doi: 10.1094/PDIS-04-11-0274

Kauss, H., Seehaus, K., Franke, R., Gilbert, S., Dietrich, R. A. and Kröger, N. (2003). Silica deposition by a strongly cationic proline-rich protein from systemically resistant cucumber plants. *Plant J.* 33: 87-95. doi: 10.1046/j.1365-313X.2003.01606.x

Kim, D., Paggi, J. M., Park, C., Bennett, C. and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37: 907-915. doi: 10.1038/s41587-019-0201-4

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27(5): 722-736. doi: 10.1101/gr.215087.116

Kovacs, K. F., Haight, R. G., McCullough, D. G., Mercader, R. J. Siegert, N. W. and Liebhold, A. M. (2010). Cost of potential emerald ash borer damages in U.S. communities, 2009-2019. *Ecol. Econ.* 69: 569-578. doi: 10.1016/j.ecolecon.2009.09.004

Kulich, I., Vojtíková, Z., Sabol, P., Ortmannová, J., Neděla, V., Tihlaříková, E. *et al*. (2018). Exocyst subunit EXO70H4 has a specific role in callose synthase secretion and silica accumulation. *Plant Physiol.* 176: 2040-2051. doi: 10.1104/pp.17.01693

Kumar, S., Soukup, M. and Elbaum, R. (2017). Silicification in grasses: variation between different cell types. *Front. Plant Sci.* 8: 438 doi: 10.3389/fpls.2017.00438

Lanning, F. C., Hopkins, T. L. and Loera, J. C. (1980). Silica and ash content and depositional patterns in tissues of mature *Zea mays* L. plants. *Ann. Bot.* 45: 549-554. doi: 10.1093/oxfordjournals.aob.a085859

Law, C. and Exley, C. (2011). New insight into silica deposition in horsetail (*Equisetum arvense*). *BMC Plant Biol.* 11: 112. doi: 10.1186/1471-2229-11-112

Lee, Y. K., Kim, G., Kim, I., Park, J., Kwak, S., Choi, G. *et al*. (2006). LONGIFOLIA1 and LONGIFOLIA2, two homologous genes, regulate

longitudinal cell elongation in *Arabidopsis*. *Development* 133: 4305-4314. doi: 10.1242/dev.02604

Lee, Y. K., Rhee, J. Y., Lee, S. H., Chung, G. C., Park, S. J., Segami, S. *et al*. (2018). Functionally redundant LNG3 and LNG4 genes regulate turgor-driven polar cell elongation through activation of XTH17 and XTH24. *Plant Mol. Biol.* 97: 23-36. doi: 10.1007/s11103-018-0722-0

Léran, S., Garg, B., Boursiac, Y., Corratgé-Faillie, C., Brachet, C., Tillard, P. *et al*. (2015). AtNPF5.5, a nitrate transporter affecting nitrogen accumulation in *Arabidopsis* embryo. *Sci. Rep.* 5: 7962. doi: 10.1038/srep07962

Leroux, O., Leroux, F., Mastroberti, A. A., Santos-Silva, F., van Loo, D., Bagniewska-Zadworna, A. *et al*. (2013). Heterogeneity of silica and glycan-epitope distribution in epidermal idioblast cell walls in *Adiantum raddianum* laminae. *Planta* 237: 1453-1464. doi: 10.1007/s00425-013-1856-6

Li, B., Xu, W., Xu, Y., Zhang, Y., Wang, T., Bai, Y. *et al*. (2010). Integrative study on proteomics, molecular physiology, and genetics reveals an accumulation of cyclophilin-like protein, TaCYP20-2, leading to an increase of Rht protein and dwarf in a novel GA-insensitive mutant (*gaid*) in wheat. *J. Proteome Res.* 9: 4242-4253. doi: 10.1021/pr100560v

Li, F., Fu, N., Zhou, J. and Wang, G. (2017). Functional characterization of (E)-β-caryophyllene synthase from lima bean and its up-regulation by spider mites and alamethicin. *J. Integr. Agric.* 16: 2231-2238. doi: 10.1016/S2095-3119(16)61593-9

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34: 3094-3100. doi: 10.1093/bioinformatics/bty191

Liebminger, E., Hüttner, S., Vavra, U., Fischl, R., Schoberer, J., Grass, J. *et al*. (2009). Class I α-mannosidases are required for n-glycan processing and root development in *Arabidopsis thaliana*. *Plant Cell* 21: 3850-3867. doi: 10.1105/tpc.109.072363

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J. *et al*. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28: 2397-2399. doi: 10.1093/bioinformatics/bts444

Liu, J., Zhu, J., Zhang, P., Han, L., Reynolds, O. L., Zeng, R. *et al*. (2017). Silicon supplementation alters the composition of herbivore induced plant

volatiles and enhances attraction of parasitoids to infested rice plants. *Front. Plant Sci.* 8: 1265 doi: 10.3389/fpls.2017.01265

Liu, Q. and Zhu, Z. (2010). Functional divergence of the NIP III subgroup proteins involved altered selective constraints and positive selection. *BMC Plant Biol.* 10: 256. doi: 10.1186/1471-2229-10-256

Love, M.I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15:550. doi: 10.1186/s13059-014-0550-8

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R. *et al*. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* 48: D265-D268. doi: 10.1093/nar/gkz991

Ma, J. F., Tamai, K., Ichii, M and Wu, G. F. (2002). A rice mutant defective in Si uptake. *Plant Physiol.* 130: 2111-2117. doi: 10.1104/pp.010348

Ma, J. F., Tamai, K., Yamaji, N., Mitani, N., Konishi, S., Katsuhara, M. *et al*. (2006). A silicon transporter in rice. *Nature* 440: 688-691. doi: 10.1038/nature04590

Ma, J. F., Yamaji, N., Mitani, N., Tamai, K., Konishi, S., Fujiwara, T. *et al*. (2007). An efflux transporter of silicon in rice. *Nature* 448: 209, 212. doi: 10.1038/nature05964

Ma, F. J., Yamaji, N. and Mitani-Ueno (2011). Transport of silicon from roots to panicles in plants. *Proc. Jpn. Acad., Ser.* B 87: 377-385. doi: 10.2183/pjab.87.377

Ma, X., Song, L., Yang, Y. and Liu, D. (2013). A gain-of-function mutation in the ROC1 gene alters plant architecture in *Arabidopsis*. *New Phytol.* 197: 751-762. doi: 10.1111/nph.12056

Manivannan, A. and Ahn, Y. (2017). Silicon regulates potential genes involved in major physiological processes in plants to combat stress. *Front. Plant Sci.* 8: 1346. doi: 10.3389/fpls.2017.01346

Mansfield, J. W., Galambos, N. and Saville, R. (2018). The use of ascospores of the dieback fungus *Hymenoscyphus fraxineus* for infection assays reveals a significant period of biotrophic interaction in penetrated ash cells. *Plant Pathol.* 67: 1354–1361doI: 10.1111/ppa.12844

Marzoug, H. N. B., Romdhane, M., Lebrihi, A., Mathieu, F., Couderc, F., Abderraba, M. *et al*. (2011). *Eucalyptus oleosa* essential oils: chemical composition and antimicrobial and antioxidant activities of the oils from different plant parts (stems, leaves, flowers and fruits). *Molecules* 16: 1695-1709. doi: 10.3390/molecules16021695

Massey, F. P. and Hartley, S. E. (2008). Physical defences wear you down: progressive and irreversible impacts of silica on insect herbivores. *J. Anim. Ecol.* 78: 281-291. doi: 10.1111/j.1365-2656.2008.01472.x

McMullan, M., Rafiqi, M., Kaithakottil, G., Glavijo, B. J., Biham, L., Orton, E. *et al*. (2018). The ash dieback invasion of Europe was founded by two genetically divergent individuals. *Nat. Ecol. Evol.* 2: 1000-1008. doi: 10.1038/s41559-018-0548-9

Miller, C. N., Harper, A. L., Trick, M., Werner, P., Waldron, K. and Bancroft, I. (2016). Elucidation of the genetic basis of variation for stem strength characteristics in bread wheat by associative transcriptomics. *BMC Genomics* 17: 500. doi: 10.1186/s12864-016-2775-2

Mitani, N., Chiba, Y., Yamaji, N. and Ma, J. F. (2009). Identification and characterisation of maize and barley LSI1-like silicon efflux transporters reveals a distinct silicon uptake system from that in rice. *Plant Cell* 21: 2133-2142. doi: 10.1105/tpc.109.067884

Mitani, N., Yamaji, N., Ago, Y., Iwasaki, K. and Ma. J. F. (2011). Isolation and functional characterization of an influx silicon transporter in two pumpkin cultivars contrasting in silicon accumulation. *Plant J.* 66: 231-240. doi: 10.1111/j.1365-313X.2011.04483.x

Mitani-Ueno, N., Yamaji, N., Zhao, F.J. and Ma, J. F. (2011). The aromatic/arginine selectivity filter of NIP aquaporins plays a critical role in substrate selectivity for silicon, boron, and arsenic. *J. Exp. Bot.* 62: 4391-4398. doi: 10.1093/jxb/err158

Montpetit, J., Vivancos, J., Mitani-Ueno, N., Yamaji, N., Rémus-Borel, W., Belzile, F. *et al*. (2012). Cloning, functional characterization and heterologous expression of *TaLsi1*, a wheat silicon transporter gene. *Plant Mol. Biol.* 79: 35-46. doi: 10.1007/s11103-012-9892-3

Mvondo-She, M. A. and Marais, D. (2019). The investigation of silicon localization and accumulation in citrus. *Plants* 8: 200. doi: 10.3390/plants8070200

Nakata, Y., Ueno, M., Kihara, J., Ichii, M., Taketa, S. & Arase, S. (2008). Rice blast disease and susceptibility to pest in a silicon uptake-deficient mutant *lsi1* of rice. *Crop Prot.* 27: 865-868. doi: 10.1016/j.cropro.2007.08.016

Nascimento-Silva, K., Roca-Castillo, L., Benlloch-González, M. and Fernández-Escobar, R. (2019). Silicon reduces the incidence of *Venturia oleaginea* (Castagne) Rossman & Crous in potted olive plants. *HortScience* 54: 1962-1966. doi: 10.21273/HORTSCI14293-19

Niemz, P., Clauss, S., Michel, F., Hänsch, D. and Hänsel, A. (2014). Physilcal and mechanical properties of common ash (*Fraxinus excelsior* L.). *Wood Research* 59: 671-682

Oh, K., Ivanchenko, M. G., White, T. J. and Lomax, T. L. (2006). The *diageotropica* gene of tomato encodes a cyclophilin: a novel player in auxin signaling. *Planta* 224: 133-144. doi: 10.1007/s00425-005-0202-z

Oyama, T., Shimura, Y. and Okada, K. (2002). The *IRE* gene encodes a protein kinase homologue and modulates root hair growth in *Arabidopsis*. *The Plant Journal* 30: 289-299. doi: 10.1046/j.1365-313X.2002.01290.x

Perry, C. C. and Lu, Y. (1992). Preparation of silicas from silicon complexes: role of cellulose in polymerisation and aggregation control. *J. Chem. Soc., Faraday Trans.* 88: 2915-2921. doi: 10.1039/FT9928802915

Perry, C. C. and Keeling-Tucker, T. (2000). Biosilicification: the role of the organic matrix in structure control. *J. Biol. Inorg. Chem.* 5: 537–550. doi: 10.1007/s007750000130

Pfeiffer, F., Gröber, C., Blank, M., Händler, K., Beyer, M., Schultze, J. L. *et al*. (2018). Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci. Rep.* 8: 10950. doi: 10.1038/s41598-018-29325-6

Popescu, A. A., Harper, A. L., Trick, M., Bancroft, I. and Huber, K. T. (2014). A novel and fast approach for population structure inference using kernel-PCA and optimization. *Genetics* 198: 1421-1431. doi: 10.1534/genetics.114.171314

Pu, X., Li, Z., Tian, Y., Gao, R., Hao, L., Hu, Y. *et al*. (2020). The honeysuckle genome provides insight into the molecular mechanism of carotenoid metabolism underlying dynamic flower coloration. *New Phytol.* 227: 930-943. doi: 10.1111/nph.16552

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL www.R-project.org/.

Reidinger, S., Ramsey, M. H. and Hartley, S. E. (2012). Rapid and accurate analyses of silicon and phosphorus in plants using a portable X-ray fluorescence spectrometer. *New Phytol.* 195: 699-706. doi: 10.1111/j.1469-8137.2012.04179.x

Romano, P. G. N., Horton, P. and Gray, J. E. (2004). The *Arabidopsis* cyclophilin gene family. *Plant Physiology* 134: 1268-1282. doi: 10.1104/pp.103.022160

Shaw, P. E. (2002). Peptidyl-prolyl isomerases: a new twist to transcription. *EMBO Rep.* 3: 521-526. doi: 10.1093/embo-reports/kvf118

Shen, W., Le, S. and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE* 11(10): e0163962. doi: 10.1371/journal.pone.0163962

Sollars, E. S. A., Harper, A. L., Kelly, L. J., Sambles, C. M., Ramirez-Gonzalez, R. H., Swarbreck, D., *et al.* (2017). Genome sequence and genetic diversity of European ash trees. *Nature* 541: 212-216. doi: 10.1038/nature20786

Song, A., Li, P., Fan, F., Li, Z. and Liang, Y. (2014). The effect of silicon on photosynthesis and expression of its relevant genes in rice (*Oryza sativa* L.) Under high-zinc stress. *PLoS ONE* 9: e113782. doi: 10.1371/journal.pone.0113782

Soukup, M., Martinka. M., Bosnić, D., Čaplovičová, M., Elbaum, R. & Lux, A. (2017). Formation of silica aggregates in sorghum root endodermis is predetermined by cell wall architecture and development. *Ann. Bot.* 120: 739-753. doi: 10.1093/aob/mcx060

Strasser, R., Schoberer, J., Jin, C., Glössl, J., Mach, L. and Steinkellner, H. (2006). Molecular cloning and characterization of *Arabidopsis thaliana* Golgi α-mannosidase II, a key enzyme in the formation of complex N-

glycans in plants. *Plant J.* 45: 789-803. doi: 10.1111/j.1365-313X.2005.02648.x

Talukdar, P., Hartley, S. E., Travis, A. J., Price, A. H. and Norton, G. J. (2019). Genotypic differences in shoot silicon concentration and the impact on grain arsenic concentration in rice. *J. Plant Nutr. Soil Sc.* 182: 265-276. doi: 10.1002/jpln.201800373

Tiller, N., Weingartner, M., Thiele, W., Maximova, E., Schöttler, M. A. and Bock, R. (2012). The plastid-specific ribosomal proteins of *Arabidopsis thaliana* can be divided into non-essential proteins and genuine ribosomal proteins. *Plant J.* 69: 302-316. doi: 10.1111/j.1365-313X.2011.04791.x

Trembath-Reichert, E., Wilson, J. P., McGlynn, S. E. and Fischer, W. W. (2015). Four hundred million years of silica biomineralization in land plants. *PNAS* 112: 5449-5454. doi: 10.1073/pnas.1500289112

Tsutsui, O., Sakamoto, R., Obayashi, M., Yamakawa, S., Handa, T., Nishio-Hamane, D. *et al*. (2016). Light and SEM observation of opal phytoliths in the mulberry leaf. *Flora* 218: 44-50. doi: 10.1016/j.flora.2015.11.006

Tyler, A. D., Mataseje, L., Urfano, C. J. Schmidt, L., Antonation, K. S., Mulvey, M. R. *et al*. (2018). Evaluation of Oxford Nanopore's MinION sequencing device for microbial whole genome sequencing applications. *Sci. Rep.* 8: 10931. doi: 10.1038/s41598-018-29334-5

Vaillancourt, B. and Buell, C. R. (2019). High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. *bioRxiv* 783159. doi: 10.1101/783159

Valenta, V., Moser, D., Kapeller, S. and Essl, F. (2017). A new forest pest in Europe: a review of Emerald ash borer (Agrilus planipennis) invasion. *J. Appl. Entomol.* 141: 507-526. doi: 10.1111/jen.12369

van Ooijen, G., Mayr, G., Kasiem, M. M. A., Albrecht, M., Cornelissen, B. J. C. and Takken, L. W. (2008). Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. *J. Exp. Bot.* 59: 1383-1397. doi: 10.1093/jxb/ern045

Vivancos, J., Labbé, C., Menzies, J. G. and Bélanger, R. R. (2014). Silicon-mediated resistance of Arabidopsis against powdery mildew involves

mechanisms other than the salicylic acid (SA)-dependent defence pathway. *Mol. Plant. Pathol.* 16: 572-582. doi: 10.1111/mpp.12213

Walker, B. J., Abeel, T. Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S. *et al*. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9: e112963

Wallace, I. S. and Roberts, D. M. (2005). Distinct transport selectivity of two structural subclasses of the nodulin-like intrinsic protein family of plant aquaglyceroporin channels. *Biochemistry* 44: 16826-16834. doi: 10.1021/bi0511888

Wang, M., Gao, L., Dong, S., Sun, Y., Shen, Q. and Guo, S. (2017). Role of silicon on plant–pathogen interactions. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00701

Winchester, B. (2005). Lysosomal metabolism of glycoproteins. *Glycobiology* 15: 1R-15R. doi: 10.1093/glycob/cwi041

Wu, T. D. and Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21: 1859-1875. doi: 10.1093/bioinformatics/bti310

Yamaji, N., Mitani, N. and Ma, J. F. (2008). A transporter regulating silicon distribution in rice shoots. *Plant Cell* 20: 1381-1389. doi: 10.1105/tpc.108.059311

Yang, B., Voiniciuc, C. Fu, L., Dieluweit, S., Klose, H. and Usadel, B. (2018). TRM 4 is essential for cellulose deposition in *Arabidopsis* seed mucilage by maintaining cortical microtubule organization and interacting with CESA 3. *New Phytol.* 221: 881-895. doi: 10.1111/nph.15442

Ye, M., Song, Y., Long, J., Wang, R., Raerson, S. R., Pan, Z. *et al*. (2013). Priming of jasmonate-mediated antiherbivore defense responses in rice by silicon. *PNAS* 110: E3631-E3639. doi: 10.1073/pnas.1305848110

Yue, X., Guo, Z., Shi, T., Song, L. and Cheng, Y. (2019). *Arabidopsis* AGC protein kinases IREH1 and IRE3 control root skewing. *J. Genet. Genomics* 46: 259-267. doi: 10.1016/j.jgg.2019.02.007

Zhao, Y., Hosoya, T., Baral, H., Hosaka, K. and Kakishima, M. (2012). *Hymenoscyphus pseudoalbidus*, the correct name for *Lambertella albida* reported from Japan. *Mycotaxon* 122: 25-41. doi: 10.5248/122.25

Zhang, W., Zhao, F., Jiang, L., Chen, C., Wu, L. and Liu, Z. (2018). Different pathogen defense strategies in *Arabidopsis*: more than pathogen recognition. *Cell* 7: 252. doi: 10.3390/cells7120252

Zhu, X. F., Shi, Y. Z., Lei, G. J., Fry, S. C., Zhang, B. C., Zhou, Y. H. *et al*. (2012). XTH31, encoding an in vitro XEH/XET-active enzyme, regulates aluminium sensitivity by modulating in vivo XET action, cell wall xyloglucan content, and aluminium binding capacity in *Arabidopsis*. *Plant Cell* 24: 4731-4747. doi: 10.1105/tpc.112.106039