s l u

Towards genomic-based breeding in Norway spruce

Linghua Zhou

Towards genomic-based breeding in Norway spruce

Linghua Zhou

Faculty of Forest Sciences
Department of Forest Genetics and Plant Physiology
Umeå

SLU

**SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES**

Acta Universitatis agriculturae Sueciae
2020:41

Cover: The reflection of Norway spruce genes
(Draw: Linghua Zhou)

Towards genomic-based breeding in Norway spruce

Abstract

Norway spruce is one of the most economically and ecologically important forest tree species in Europe. The central obstacle to Norway spruce breeding is the length of the breeding cycle, which takes 20 years or more to be completed. By utilizing genomic-based breeding, breeding cycle length can be reduced, and accuracy of selection can be improved. This thesis evaluates the potential of marker-assisted selection and genomic selection in Norway spruce breeding.

Breeding values from 517 independent plus trees and 178101 Single Nucleotide Polymorphisms (SNPs) generated with the exome capture approach on those same trees, were used to conduct Genome-Wide Association Studies (GWAS) for 17 solid wood quality traits and 15 wood tracheid properties in Norway spruce. Together, 52 significant SNPs from 39 candidate genes and 31 SNPs from 26 candidate genes were identified. We also found 11 significant SNPs associated with resistance to *Heterobasidium parviporum* in a population of 466 trees. GWAS was used as a tool to detect genes determining fungal community composition in dormant buds based on data from 478 plus trees. Predictive Ability (PA) of Genomic selection (GS) was evaluated in 484 progeny trees from 62 half-sib families for solid wood quality traits measured with12mm increment wood cores and standing trees. Results from the genomic-based method are similar to those from the pedigree-based method.

In this thesis, the genetic information rendered by GWAS is insufficient to conduct efficient marker-assisted selection (MAS), however it has advanced our knowledge of the genetic architecture of traits of economic and ecological value, as well as their genetic correlations. On the other hand, GS is considered as a powerful alternative to genomic-based breeding in Norway spruce.

Keywords: GWAS, GS, wood quality traits, tracheid properties, Heterobasidium resistance, fungal communities

Author's address: Linghua Zhou, Swedish University of Agricultural Sciences, Department of Forest Genetics and Plant Physiology, Umeå, Sweden

Preface

2020 is a special year. My 2020 started in my favourite city, my hometown, due to an unexpected trip. Then the coronavirus pandemic spread all over the world. People were forced to stay at home to control the infection. No medicine and no vaccine. Human beings seem so tiny in the face of such a disease. Not only did the medical resources in the world reach the edge, but also our knowledge. Likewise, human feel so little when standing in the middle of the forest surrounded by tall trees. Height, as many other tree properties, is controlled by environment and genetics. The tree breeders' dream is to produce the best tree with the highest yearly gain in any environment, and quantitative genetics is the tool that breeders apply to achieve that dream, as well as caring for the impact of their actions on biodiversity and the environment. Long ago scientists and researchers could only separate the effect between environment and genetics roughly according to limited information, but that early approach soon reached its limit. In the 20th century, with the development of genomic technology, the gate to the genomic world opened. Now, huge databases filled with genomic data have been produced for human, animals, small plants and big trees. The genome for several species of conifer has been sequenced and assembled, even though they are enormous and complex, Norway spruce being one of them. So, we are one more step closer to the breeders' dream. Hopefully, we are one more step closer to defeating the coronavirus too.

Dedication

To people I love.

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. **Zhou L.**, Chen Z.-Q., Lundqvist S-O., Olsson L., Grahn T., Karlsson B., Wu H., García-Gil M.R. (2019). Genetic analysis of wood quality traits in Norway spruce open-pollinated progenies and their parent plus trees at clonal archives and the evaluation of phenotypic selection of plus tree. Canadian Journal of Forest Research, 2019, 49(7): 810-818

II. Baison, J., Vidalis, A., **Zhou, L.**, Chen, Z.-Q., Li, Z., Sillanpää, M.J., Bernhardsson, C., Scofield, D., Forsberg, N., Grahn, T., Olsson, L., Karlsson, B., Wu, H., Ingvarsson, P.K., Lundqvist, S-O., Niittylä, T. and García-Gil, M.R. (2019). Genome-wide association study identified novel candidate loci affecting wood formation in Norway spruce. Plant J, 100: 83-100.

III. Elfstrand M., Baison J., Lundén K., **Zhou L.**, Vos I., Capador-Barreto H.D., Stein Åslund M., Chen Z., Chaudhary R., Olson Å., Wu H.X., Karlsson B., Stenlid J., García-Gil M.R. (2020). Association genetics identifies a specifically regulated Norway spruce laccase gene, PaLAC5, linked to Heterobasidion parviporum resistance. Plant Cell Environ. 2020; 1– 13.

IV. Elfstrand M., **Zhou L.**, Baison J., Olson Å., Lundén K., Karlsson B., Wu H., Stenlid J., García-Gil M.R. (2020). Genotypic variation in Norway spruce correlates to fungal communities in vegetative buds. Mol Ecol. 2020; 29: 199– 213.

V. **Zhou, L.**, Chen, Z., Olsson, L., Grahn T., Karlsson B., Wu H., Lundqvist, S-O., García-Gil M.R. (2020). Effect of number of annual rings and tree ages on genomic

predictive ability for solid wood properties of Norway spruce. BMC
Genomics 21, 323

VI. Baison J., **Zhou L.**, Forsberg N., Mörling T., Grahn T., Olsson L., Karlsson B., Wu
H., Mellerowicz E., Lundqvist S-O., García-Gil M.R. (2020). Genetic control of
tracheid properties in Norway spruce wood. Scientific Reports (submitted)

Papers I-V are reproduced with the permission of the publishers.

List of tables

List of figures

Abbreviations

ABLUP	A-matrix Best Linear Unbiased Predictor
EBVs	Estimated Breeding Values
EW	Earlywood
GBLUP	G-matrix Best Linear Unbiased Predictor
GEBV	Genomic Estimated Breeding Value
GS	Genomic Selection
GWAS	Genome-Wide Associate Study
h^2	Narrow sense heritability
LD	Linkage Disequilibrium
LG	Linkage Group
LL	Lesion Length
LW	Latewood
MAS	Marker-Assisted Selection
MFA	MicroFiber Angle
MOE	Modulus Of Elasticity
OTU	Operational Taxonomic Unit
PA	Predictive Ability
PCA	Principle Component Analysis
QTL	Quantitative Trait Loci
SNP	Single Nucleotide Polymorphism

SWG   Sapwood fungal growth

TW   Transition wood

1. Introduction

Forest trees are the group of plant species most likely to benefit from the use of genomic information in the context of breeding (Plomion et al., 2016). Early genomic-based evaluations are expected to increase the precision of selection and shorten the breeding cycle for forest trees.

1.1 Development of molecular marker-based breeding in forest trees

1.1.1 DNA-based molecular markers

All living organism are made up of cells programmed by a molecule called DNA. This molecule is made up of a long chain of nitrogen-containing bases. There are four different bases – adenine (A), cytosine (C), guanine (G) and thymine (T). DNA is transmitted according to Mendelian laws of inheritance from one generation to the next.

DNA-based molecular markers, or loci, are regions defined along the DNA. Each molecular marker is described by the length or nucleotide composition of its nucleotide sequence, which can be as simple as a single nucleotide. Either way, a molecular marker aims to define the level of polymorphism at a given locus, which can be located at genic or non-genic regions. A molecular marker targeting a single nucleotide locus is called Single Nucleotide Polymorphism (SNP), and could potentially be as abundant as the number of base-pairs along an organism´s DNA molecule. With the advent of high-throughput sequencing, SNPs have become the most cost-effective molecular markers.

1.1.2   DNA-based molecular marker applications in forest genetics

The first molecular marker-based attempts in the context of forest tree breeding began in the 1990s (Emebiri et al., 1997; Groover et al., 1994). Even at that time, many studies succeeded in identifying quantitative trait loci (QTL) in agronomic crops (McCough and Doerge, 1995; Procunier et al., 1997; Young et al., 1992). In forest genetics, the early QTL studies soon provided evidence that a handful of QTLs of minor effect would not assist in early selection in complex traits (Strauss et al., 1992). The lack of genetic resolution of those early studies was the result of the high cost of genotyping large number of trees at the required high molecular marker density. It is important to note that the sample sizes and molecular marker densities in the early QTL studies were insufficient considering that conifers are characterized by large, complex and highly polymorphic genomes, in which linkage disequilibrium (LD) decays rapidly (Neale and Savolainen, 2004).

In the early 2000s, with the advent and rapid development of DNA sequencing technology, screening of a larger number of SNPs became possible. More recently, new technological developments have resulted in affordable SNP screening at the whole-genome level of large sample sizes. In forest tree genetics, SNP markers have been applied to statistically reconstruct linkage groups (LGs) (i.e., a representation of the gene composition of a chromosome where genes are arranged by the level of linkage among them) (Bernhardsson et al., 2019; Pavy et al., 2017), to identify the main genes (as QTLs) underlying a trait (Eckert et al., 2012; Lamara et al., 2016; Parchman et al., 2012), and more recently to improve efficiency of selection in breeding programs (Beaulieu et al., 2014b; Chen et al., 2018a; Isik et al., 2016; Lenz et al., 2017).

1.1.3   Family-based QTL mapping

QTL mapping, also called linkage mapping, requires association across loci which can be generated by conducting bi-parental crosses (families). In those crosses, co-segregation between alleles at marker loci and phenotypic trait allows the identification of QTLs underlying a trait (Laidò et al., 2014). In plants, several main-effect QTLs have been identified for yield (Borner et al., 2002; Kumar et al., 2007; Maccaferri et al., 2008; Peng et al., 2003), grain protein content (Olmos et al., 2003; Sun et al., 2010), disease resistance (Marone et al., 2013; Marone et al., 2009; Navabi et al., 2005;

Todorovska et al., 2009), and flowering time (Båga et al., 2009; Hanocq et al., 2004; Panio et al., 2013).

In forest trees - especially coniferous species - QTL analyses have failed to identify major QTLs for complex traits with the same effectiveness as for crop species (Neale and Kremer, 2011). Instead, few tens of QTLs of minor effect are often detected (Hall et al., 2016). This is mainly a consequence of the complexity and size of the conifer genome, and the rapid decay of LD, which can only be compensated for by experimental designs that involve tens of thousands of trees from one or less often from several families. However, in reality, most QTL studies have been conducted on a few or only one full-sib families, with small family sizes - in most cases in the range of 100s of trees (Gion et al., 2011; Jermstad et al., 2001; Pot et al., 2006). Furthermore, QTLs identified in a single full-sib family cannot be extrapolated to other families where other alleles, other epistatic interactions (ie., interaction across loci), or simply change in allele frequency (e.g., a locus becomes non-polymorphic) may define those same loci. Nor is it possible - in most cases - to extrapolate QTL effects across environments (Dillen et al. 2008; Freeman et al. 2013; Gion et al. 2011; Novaes et al. 2009; Rae et al. 2008; Thumma et al. 2010). These represent major limitations considering that breeding is conducted with information that involves hundreds of families (Skogforsk, 2011). Finally, estimated QTL effects are not only family- and environment-dependent but are often over-estimated (the so-called Beavis effect) mostly due to the limited statistical power of the experimental designs (Beavis and Paterson, 1998; Hall et al., 2016).

1.1.4   Association mapping

Association mapping, also known as LD mapping, is fundamentally similar to linkage mapping, but in this case the focus changes from families to populations (Laidò et al., 2014). LD mapping relies on historical recombination events accumulated over generations, and the presence of a larger number of alleles where it is more likely that a significant marker is physically near to the causal gene, or within the gene itself (Ingvarsson et al., 2016). In this respect, if a reasonable proportion of genetic variation is explained by any molecular marker(s) it could potentially be applied to conduct early Marker-Assisted Selection (MAS) (O'Connor et al., 2020). The fact that LD mapping does not rely on controlled crosses represents a major advantage as it overcomes one of the main limitations of family-based

QTL analyses. Instead, any existing forest can be the target of a LD mapping study.

When LD mapping is conducted at the whole genome level, it is known as Genome-Wide Association Mapping (GWAS) (Risch and Merikangas, 1996). In combination with high-throughput sequencing technology, GWAS is a powerful tool for the identification of causative gene(s) underlying complex traits in model and non-model forest tree species (Alqudah et al., 2020).

In conifers, a high density of molecular markers is required to dissect a complex trait, due to the large genome size and the rapid decay of LD (Neale and Savolainen, 2004; Neale and Kremer, 2011). GWAS has already yielded important insights into the genetic basis of complex quantitative traits of conifer species, focusing on wood composition in Loblolly Pine (Eckert et al., 2012) and wood property traits in white spruce (Lamara et al., 2016), Cedar (Hiraoka et al., 2018). These studies were important for tree breeding programs and helped researchers to understand many aspects of the ecological genetics of conifers (McKown et al., 2014b).

1.1.5   Genomic selection

Genome-wide prediction, or genomic selection (GS), relies on simultaneously estimating the effects of many thousands of SNPs across the whole genome to estimate the genetic merit, and genomic estimated breeding value (GEBV), of individuals based purely on genomic information (Meuwissen et al., 2001). GS therefore avoids the possibility of missing a substantial portion of genetic variance associated to large numbers of loci of minor effect (Bhat et al., 2016).

Conventional methods rely on pedigree information (ABLUP, A-matrix best linear unbiased predictor) to estimate the genetic merit of an individual. Instead, GS relies on molecular markers to estimate the realized genomic relationships between trees (GBLUP, G-matrix best linear unbiased predictor). This approach can accurately capture relatedness at within-family levels, thereby increasing the precision of evaluations (Plomion et al., 2016).

Accuracy of genomic prediction in forest tree species has been mainly tested in cross-validation designs, where full-sib and/or half-sib progenies within a single generation are subdivided into training and validation sets (Beaulieu et al., 2014b; Grattapaglia and Resende, 2011; Resende Jr et al., 2012; Resende et al., 2012). The training set is phenotyped (i.e., trait being measured) and genotyped (i.e., SNPs being scored) with the purpose of

building the genomic prediction model, whereas the validation set is only genotyped and used to evaluate the precision of the genomic model. Alternatively, other experimental designs are possible for model training and validation. For example, a study of *Pinus pinaster* based on a three generation pedigree (G0, G1, and G2) showed encouraging results based on intergeneration genomic prediction, both by mixing parents and progeny in the same training set (Isik et al., 2016) and by using G0 and G1 individuals to predict the G2 generation (Bartholomé et al., 2016).

Overall, model accuracy of GS has been reported to increase with larger training-to-validation set ratios (Chen et al., 2018a; Tan et al., 2017; Zapata-Valenzuela et al., 2013), while the level of relatedness between the two sets is considered a prominent factor (Beaulieu et al., 2014b; Chen et al., 2018a; Lenz et al., 2017; Suontama et al., 2019; Tan et al., 2017). Predictions have also been shown to be more accurate when selection is carried out across similar environments and tree-ages (El-Dien et al., 2015).

GS has been shown to increase the genetic gain per year compared to conventional breeding in many forest tree species (Beaulieu et al., 2014b; El-Dien et al., 2015; Isik et al., 2016; Zapata-Valenzuela et al., 2013). In eucalypts and poplars, the GS breeding strategy abolished the need for tree testing in progeny trials. It also shortened the time and costs involved in the clonal testing phase by reducing the number of selected trees that are evaluated as clones (Grattapaglia, 2017). In conifers, GS combined with somatic embryogenesis (SE) allowed pre-selection of zygotic embryos based on their GEBVs, which greatly improved the efficiency of propagation (Resende Jr et al., 2012).

1.2   Norway spruce breeding in Sweden

In Sweden, breeding of *Picea abies* (L.) H. Karst. (Norway spruce) and *Pinus sylvestris* L. (Scots pine) began in the 1940s and 1950s with phenotypic selection of plus trees and establishment of the first-generation seed orchards (Karlsson and Rosvall, 1993). The Norway spruce breeding program follows the multiple breeding population system (MBPS) with 20 partially overlapping, latitudinally distributed breeding populations (also called breeding zones), from which 50 trees are selected (Rosvall, 2019). These trees serve as a founder stock for long-term breeding and gene conservation programs (Androsiuk et al., 2013). In the 1980s, additional plus trees selected

from commercial forest nurseries, together with a selection of the best progeny tested plus trees from the previous selection resulted in the second round of seed orchards. At the beginning of the 21st century, a third round of seed orchards was established with the genetically best trees, selected based on results from field tests.

In each breeding cycle, a number of trees are either control-fertilized to generate full-sib progenies, or allowed open-pollination that results in half-sib progenies. Next, progenies are tested in progeny trials, where the experimental design aims for sound statistical analysis. Progeny trials are replicated across a range of locations to improve the accuracy of the genetic gain estimations (i.e., obtain the most accurate tree rank). Phenotypically or genotypically Estimated Breeding Values (EBVs) are then utilized to conduct backward selection. In other words, selection of the best plus trees based on the performance of the progeny tested. Alternatively, EBVs can be used to select the best progeny, a process called forward selection. Either way, the best trees will be deployed as seed donors in seed orchards and after reaching maturity (generation time), crossed again to start a second breeding cycle.

The major obstacle for improving wood properties is the time necessary to complete a breeding cycle, where the generation time is the major bottleneck (Markussen et al., 2004). In Sweden, the Norway spruce breeding cycle needs more than 20 years to be completed (Karlsson and Rosvall, 1993). After selection of the donor parents (e.g., plus trees), the cycle continues with a phase of breeding (crossing), followed by progeny propagation in the nursery. Seedlings are then established in progeny trials and phenotypically evaluated after certain age which depends on the trait. The step of phenotypic evaluation and genetic assessment of the mother (backward selection) or progeny (forward selection) genetic values can last more than 15 years. A summary of the breeding cycle is presented in Figure 1.
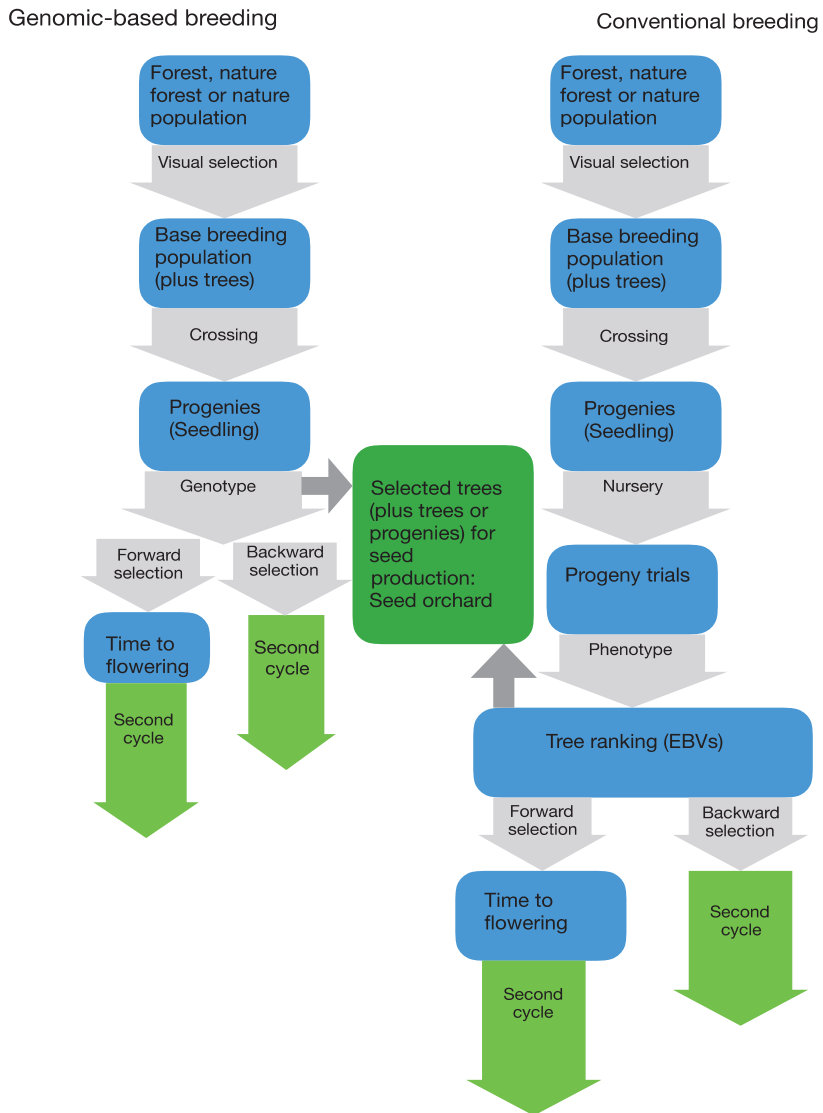
Figure 1. Comparison between conventional breeding and genomic‑ based breeding.

With genomic-based breeding, the field progeny test can be replaced by statistical genomic models involving only genomic data (SNP genotypes),

and accurately select seedlings within two years of crossing (Figure 1). This is feasible in the context of backward selection, where already mature plus trees are those being selected. However, in the case of forward selection, genomic selection of progeny, as early as two years after crossing, does not overcome the need to wait 20 years of generation time until those selected progeny reach maturity and can be crossed to start a second round of the breeding cycle. This strongly indicates that GS can only be efficiently implemented in a forward selection breeding strategy if early flowering can be induced.

1.2.1 Breeding objectives

In forest trees, especially at northern latitudes, rotation times are between nine and sixteen years for fast growing species, such as poplars. However, for slow growing species, like Norway spruce, the rotation time is around a hundred years. Therefore, the breeding goals need to remain very general and relevant over a long-term period of time. The main breeding objectives for Norway spruce are survival, resilience to abiotic and biotic stresses such as frost and pathogens, together with growth, and -more recently- wood properties (Rosvall et al., 2019).

Wood quality traits are key determinants of the tree's economic value because they determine the quality of a variety of end-products, such as structural boards, pulp and paper products, and furniture (Zobel and van Buijtenen, 1989). Wood stiffness (generally expressed in terms of the modulus of elasticity: MOE) is one of the most important traits for construction timber and is mainly determined by measuring wood density and microfiber angle (MFA) (Baltunis et al., 2007; Chen et al., 2014; Lenz et al., 2011). In addition, wall thickness, radial fiber width, tangential fiber width, and fiber coarseness are important traits contributing to overall fiber quality (Scallan and Green, 2007) which is closely related to pulp and paper properties.

Heterobasidion parviporum (*H. parviporum*) is the causal agent of annosum root rot, a common and serious fungal disease found in conifer forests of the Northern Hemisphere (Asiegbu et al., 2005), resulting in an estimated financial loss of approximately 475 million SEK (€54 million) annually (Hellgren and Stenlid, 1995). Analysis of genetic variation for resistance to *H. parviporum* in Norway spruce revealed genetic inheritance of the property and, therefore, the potential for improvement through breeding (Chen et al., 2018b).

2. Objectives

The main objectives of the research in this thesis, were to investigate the efficiency of genomic-based methods when used to assist the breeding of Norway spruce. The specific questions addressed were:

- ✧ Is selection of plus trees as reliable as progeny-based selection? (paper I)
- ✧ Are there QTLs of large effect for wood properties and resistance to *H. parviporum* (paper II, paper III & paper VI) to assist breeding?
- ✧ Is Genomic Selection a feasible alternative method to genomic-based breeding for Norway spruce? (paper V)
- ✧ Is there correlation between genotypic variation and fungal communities in vegetative buds? (paper IV)

3. Materials and methods

This chapter provides an overview of the materials and methods associated with the six papers. For more detailed information, see the corresponding papers.

3.1 Tree materials

A two-generation Norway spruce pedigree was the core material of this thesis work. The maternal generation consists of 524 plus-tree clones (genotypes) located at two different clonal archives, one in Ekebo and the other in Maltesholm, both in southern Sweden. The first-generation progeny consists of 524 open-pollinated (half-sib) families replicated in two different progeny trials, S21F9021146 aka F1146 (Höreda, Eksjö, Sweden) and S21F9021147 aka F1147 (Erikstorp, Tollarp, Sweden), which were established in 1990. We randomly selected six progenies per family and per progeny trial. In total, 12 progenies per half-sib family were included in the study. The plants are spaced on a grid of 1.4m×1.4m for the progeny trials. In the clonal archives, the original spacing was 3m×0.5m, though this was gradually increased following thinning. At the time of establishment, there were originally 10 ramets (identical genotypes) grafted per clone at the clonal archive, while at the time of sampling only three ramets remained in the archive. The plus trees in the archives were grafted on rootstocks at the seedling stage. As part of their maintenance, both clone archives were topped in the autumn of 2007 at age 23, when a large seed crop was harvested. Thinning of the Ekebo clonal archive and parts of the Maltesholm archive were carried out for the first time in the late 1990s, and the most recent thinning was conducted in the autumn of 2009 at age 25.

In Paper I, II and VI, data from the two-generation trees were utilized, in paper III and IV, data from a subset of the plus trees (466 and 478

respectively) were used, and in paper V, 484 progenies from 62 families were utilized.

3.2 Phenotype

3.2.1 Wood quality traits

Increment cores of 12 mm diameter were sampled in 2010 from one ramet per plus-tree and from all 12 progeny trees at age 21-years old (Chen et al. 2014). Each core was screened using a SilviScan instrument. All three plus-tree ramets and same progenies above mentioned were also scored for pilodyn penetration depth using a Pilodyn 6J Forest (PROCEQ, Zurich, Switzerland) and acoustic velocity using a Hitman ST300 (Fiber-gen, Christchurch; New Zealand) at ages 22 and 24, respectively (Chen et al. 2015). All traits are summarized in Table 1.

Table 1. Wood quality traits involved in this thesis

Measurement method	Trait	Abbreviation	Unit
Increment core-based (SliviScan)	Wood density	WD	kg/m^3
	Ring width	RW	μm
	Modulus of elasticity	MOE	GPa
	Microfiber angel	MFA	Degree
	Number of cells	NC	
	Radial tracheid width	RTW	μm
	Tangential tracheid width	TW	μm
	Wall thickness	WT	μm
	Coarseness	CO	mg/m
Standing tree-based	Pilodyn	Pilo	mm
	Velocity	AV	(km/s)^2
	Modulus of elasticity	MOE$_{ind}$	GPa

3.2.1.1 SliviScan

The SilviScan is an instrument used for efficient measurement of multiple solid wood properties. On increment cores, it delivers data from pith to bark at high spatial resolution (Evans 1994; Evans 2006). Each increment core was turned into 2mm sanded strips representing all the rings from pith to bark and automatically scanned for radial variations in cross-sectional

tracheid widths with X-ray transmission and diffraction, combined with light image analysis. From these data, the locations of the annual rings were first identified, followed by the identification of three compartments: earlywood (EW), transition wood (TW) and latewood (LW) using the "20-80 density" definition (Lundqvist et al., 2018). The measurement was conducted at Innventia, now RISE Bioeconomy, Stockholm, Sweden.

3.2.1.2   Indirect measurements

In addition, measurements of the three solid wood traits (density, MFA and MOE) were also obtained with Pilodyn and Hitman instruments conducted directly on standing trees without removing the bark, at ages 22 and 24 years. The Pilodyn measures penetration depth by pressing a needle into the stem. Depth of penetration is inversely correlated with wood density. The Hitman measures the velocity of sound waves in the stem, which correlates with MFA (Downes et al., 2002; Lenz et al., 2013). MOE_{ind} is related to wood density and velocity of sound (Haines and Leban, 1997; Knowles et al., 2004; Lindström et al., 2002) and can therefore be estimated by combining the Pilodyn and Velocity data, using the formula:

$$MOE_{ind} = (1/Pilo) \times 10 \times AV^2$$

where Pilo is the Pilodyn penetration depth (mm) and AV is the velocity of the wave through the material (km/s), which has a high inverse correlation with MFA (Chen et al 2015).

3.2.2   Disease resistance

Four traits related to disease resistance were included in the associate study in paper II. These were: diameter of the tree at the inoculation site (D), length of the necrotic lesion in the phloem and inner bark (LL), fungal growth in the sapwood of progeny (SWG) and tree vitality (Vitality). The induced defense response, measured by the LL parameter, was estimated by measuring the discernible lesion spread upwards and downwards from the edge of the inoculation point on the inside of the bark. SWG was estimated using established protocols (Arnerup et al., 2010; Stenlid and Swedjemark, 1988). Vitality of the progeny was scored on a scale of 1-3, where 1 was given for fully vital, and a score of 3 given to plants showing a pronounced loss of vitality.

3.3   Genotype

Genomic DNA was extracted from buds, or needles when buds were not available. This DNA was submitted to RAPiD Genomics (USA), where DNA library preparation and capture sequencing were performed. Sequence capture was carried out using the 40 018 diploid probes designed and evaluated for Norway spruce (Vidalis et al., 2018). The Illumina sequencing compatible libraries were amplified with 14 cycles of polymerase chain reaction (PCR) and the probes were then hybridized to make a pool comprising 500 ng of 8 equimolar combined libraries following Agilent's SureSelect Target Enrichment System (Agilent Technologies, https://www.agilent.com/). These enriched libraries were then sequenced using an Illumina HiSeq 2500 instrument (San Diego, USA) on the 2×100 bp sequencing mode. Raw reads were mapped against the *P. abies* reference genome v.1.0 using bwa - mem (Li, 2013). Samtools v.1.2 (Li et al., 2009) and Picard (http://broadinstitute.github.io/picard) were used for sorting and marking of PCR duplicates. Variant calling was performed using gatk haplotypecaller v.3.6 (Van der Auwera et al., 2013) in gVCF output format.

   The Variant Quality Score Recalibration (VQSR) method was performed to avoid the use of hard filtering for exome/sequence capture data. Two datasets were created for the VQSR analysis, a training subset and an input file. The training dataset was derived from the Norway spruce genetic mapping population showing expected segregation patterns (Bernhardsson et al., 2019) and assigned a prior value of 15.0. The input file was derived from the raw sequence data using gatk with the following parameters: extended probe coordinates by +100 excluding INDELS, excluding LowQual sites, and keeping only bi - allelic sites. The following annotation parameters QualByDepth, MappingQuality and BaseQRankSum, with tranches 100, 99.9, 99.0 and 90.0 were then applied for the determination of the good versus bad variant annotation profiles. After obtaining the variant annotation profiles, the recalibration was then applied to filter the raw variants.

3.4 Method

3.4.1 Mixed model

Linear mixed model was the core statistical model used in this thesis. The general model equation is:

$$y = Xb + Zu + e. \qquad \text{equation 1}$$

with $E(y) = Xb$ and $Var(y) = V = ZGZ' + R$; where y is the vector of each individual tree observation, b is the vector of the fixed effects, u is the vector of random effects and e is the vector of residuals. X and Z are incidence matrices of fixed effect (b) and random effect (u), respectively. The u and e are assumed to be independent and normally distributed as $u \sim (0, \sigma_u^2 A)$, $e \sim (0, I\sigma_e^2)$, where A is the additive genetic relationship matrix (pedigree based), or G in the case of marker-based analysis, and I is the identity matrix.

In equation 1, fixed and random effects were obtained by solving the mixed model equation (White and Hodge, 2013):

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+I\alpha \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \qquad \text{equation 2}$$

where $\hat{b}$ and $\hat{u}$ are the estimation of fixed effect and random effect respectively. I is the identity matrix with dimensions equal to the number of mothers. The scalar α is defined as $\alpha = \sigma_e^2/\sigma_A^2$, where σ_e^2 is the residual variance, σ_A^2 is the additive genetic variance which can be calculated from pedigree or marker information. $\hat{u}$ is the best linear unbiased predictions (BLUP) for additive genetic effect which is also called breeding value (BV).

3.4.2 Heritability

The individual-tree narrow-sense heritability (h^2) is the additive genetic portion of the phenotypic variance and defined as:

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2} \qquad \text{equation 3}$$

where σ_A^2 is the additive genetic variance and σ_P^2 is the phenotypic variance (Lynch and Walsh, 1998).

With two generation phenotypic data, narrow sense heritability can also be calculated by parent-offspring resemblance. The degree of resemblance is expressed as the regression of offspring on parents as in equation 4:

$$Y = \beta_0 + \beta_1 X \qquad\qquad\qquad \text{equation 4}$$

where Y is the phenotypic value for the offspring and X is the phenotypic value for the mother and since the genetic covariance between parents and offspring is equal to $0.5\sigma_A^2$ (Falconer and Mackay, 1996), β_1 can be defined as follows,

$$\beta_1 = \frac{Cov(X,Y)}{VarX} = \frac{\frac{1}{2}\sigma_A^2}{\sigma_P^2} \qquad\qquad \text{equation 5}$$

From the slope of the regression, h^2 (see equation 3) can be estimated as

$$h^2 = 2\beta_1 \qquad\qquad\qquad \text{equation 6}$$

The other method for h^2 estimation is based only on half-sib family progenies. Since the additive genetic relationship between half-sib offsprings is $0.25\sigma_A^2$ (i.e., a quarter of their alleles in common on average) and $\sigma_A^2 = 4\times$ between families variance, we can re-write equation 3 as follows,

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2} = \frac{4\times between\ families\ variance}{\sigma_P^2} \qquad \text{equation 7}$$

3.4.3   Trait association mapping

3.4.3.1   Functional model

Instead of using a single value for each trait, in paper II we introduced latent traits which were obtained by modeling the data across annual rings (i.e., from pith to bark) with the following linear spline function,

$$y(t) = \beta_0 + \beta_1 t + \beta_2 (t - K_1) + \beta_3 (t - K_2) \qquad\qquad \text{equation 8}$$

where $y(t)$ is the function of trait based on time t (annual ring), K_i ($i = 1,2$; $K_1 < K_2$) are defined as knots, and $(t - K_i)_+ = (t - K_i)$_ if $t > K_i$ ($K_i > 0$; $i = 1,2$), otherwise is equal to zero. An example for trait ring width was presented in figure 2.

In equation 8 the intercept β_0, slope parameters β_1, β_2 (at Knot 1 (K_1)) and β_3 (at Knot 2 (K_2)) are estimated by standard least squares (Ruppert et al., 2003). These four estimates were used as the latent trait in the subsequent QTL analysis. In this way, we succeeded in capturing not only the value of that trait, but also the trend of the trait across annual rings.
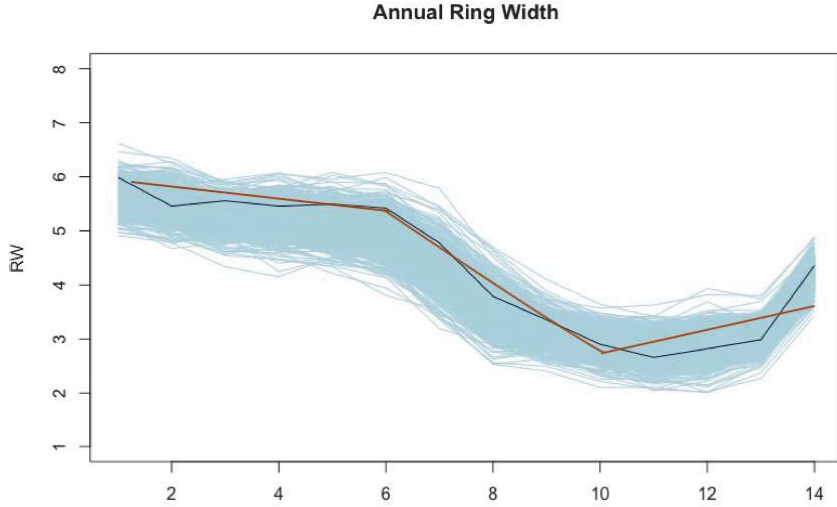
Figure 2. Trait ring width across 14 consecutive cambial ages. Light blue lines are individual samples, black line is the mean value and red line is the spline function used to map the curve.

LASSO (Least Absolute Shrinkage and Selection Operator) is a regression method that involves penalizing the absolute size of the regression coefficients. It allows handling highly dimensional cases with p>>n (i.e., the number of SNPs is much larger than the number of individuals) and selecting only a single representative SNP from the group of highly dependent SNPs.

In this thesis, we utilized the LASSO model as below:

$$\min_{(\alpha_0, \alpha_j)} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \alpha_0 - \sum_{j=1}^{p} x_{ij} \alpha_j)^2 + \lambda \sum_{j=1}^{p} |\alpha_j| \qquad \text{equation 9}$$

where y_i is the phenotypic value of an individual i ($i = 1,…,n$; n is the total number of individuals) for the latent trait β_0, β_1, β_2 or β_3, α_0 is the population mean parameter, x_{ij} is the genotypic value of individual i and marker j coded as 0, 1 and 2 for three marker genotypes AA, AB and BB, respectively. α_j is the effect of marker j ($j = 1,…,p$; p is the total number of markers), and λ (>0) is a shrinkage tuning parameter. The penalty term is able to shrink the additive effects of some of the markers exactly to zero and select the subset of markers with effect on the trait to be used by the model. By tuning the parameter λ the degree of shrinkage can be determined, as well as the number of markers having nonzero effects. Cross-validation is used to decide an optimal value for λ. Stability selection probability (SSP) for each selected

SNP was applied as a way to control the false discovery rate and determine significant SNPs (Gao et al., 2014; Li and Sillanpää, 2015).

3.4.4 Genomic prediction

In this thesis, four genomic-based methods were used to estimate genomic breeding values: Genomic Best Linear Unbiased Predictions (GBLUP) (VanRaden, 2008), Random Regression-best Linear Unbiased Predictions (rrBLUP) (Whittaker et al., 2000), BayesB (Hayes and Goddard, 2001), and Reproducing Kernel Hilbert Space (RKHS). rrBLUP applies a shrinkage parameter, lamda, to a mixed model and assumes that all markers have a common variance. In BayesB, the assumption of common variance across marker effects is relaxed by adding more flexibility to the model. RKHS does not assume linearity so it can potentially capture non-additive relationships (Heslot et al., 2012).

4. Main results and discussion

4.1 Cost-effective methods to select outstanding genotypes for solid wood properties

In paper I we evaluated the potential of ranking and selecting outstanding genotypes for solid-wood quality directly from plus-tree clonal archives. It is an alternative to backward selection based on EBVs, which requires establishment and evaluation of large numbers of half-sib progeny trials. The estimated correlations between half-sib progeny's BVs and the plus tree's phenotypic values were 0.29, 0.13 and 0.23 for Pilodyn, Velocity and MOE(ind), respectively. When using data from all three plus-tree ramets, the correlation increased to 0.32, 0.15 and 0.28 for the same traits. The narrow spacing between plus trees which resulted in strong competition may have contributed to the discrepancies between these two plantations.

Based on indirect measurements, parent-offspring h^2 estimates were higher as compared to estimates based solely on half-sib progeny data when all three ramets were included in the estimations (Table 2).

Table 2. Heritability and repeatability estimates based on measurements of wood density, MFA and MOE with SilviScan from increment cores and based on Pilodyn and Hitman (velocity).

Method	SliviScan			Indirect methods		
	Density	MFA	MOE	Pilodyn	Velocity	MOE$_{ind}$
PO* regression with_1 ramet	0.35 (±0.16)	0.15 (±0.16)	0.28 (±0.16)	0.27 (±0.15)	0.13 (±0.15)	0.13 (±0.15)
PO regression with 3 ramets	N/A	N/A	N/A	0.41 (±0.15)	0.29 (±0.15)	0.30 (±0.15)

Half-sib correlation	0.43 (±0.09)	0.29 (±0.08)	0.38 (±0.08)	0.31 (±0.08)	0.20 (±0.08)	0.28 (±0.08)
Repeatability	N/A	N/A	N/A	0.52 (±0.06)	0.30 (±0.04)	0.45 (±0.05)

* PO: Parent offspring

   This indicates that multiple copies of ramets are critical to increase the accuracy of the genetic parameter estimates for selection of outstanding plus-tree genotypes. In conclusion, backward selection, whether based on offspring data alone or a combination of offspring and clonal archive data would be most effective for wood density, and least effective for MFA, whereas MOE showed an intermediate behavior.

   Another alternative would be to select plus trees based on measurements directly obtained from the plus trees in the archive. Repeatability estimates for wood properties measured with the two indirect methods were higher compared to any h^2 estimate, indicating that selection of the best plus-tree clones could be a highly cost- and time-effective alternative to progeny trial testing.

4.2   GWAS is a genomic-based method for the dissection of complex traits

4.2.1   Wood quality traits

In paper II and VI, we identified 83 significant SNPs from 65 candidate genes related to solid wood traits (wood density, ring width, MOE, number of cells, early wood late wood proportion, mass index) and tracheid properties (microfiber angel, radiate fiber width, tangential fiber width, coarseness, wall thickness).

   Figure 3 shows a summary of the significant SNPs distributed across LGs which were produced according to the ultra - dense genetic map (Bernhardsson et al., 2019). X-axis is the position of QTL and y-axis is the proportion of variance explained (PVE).
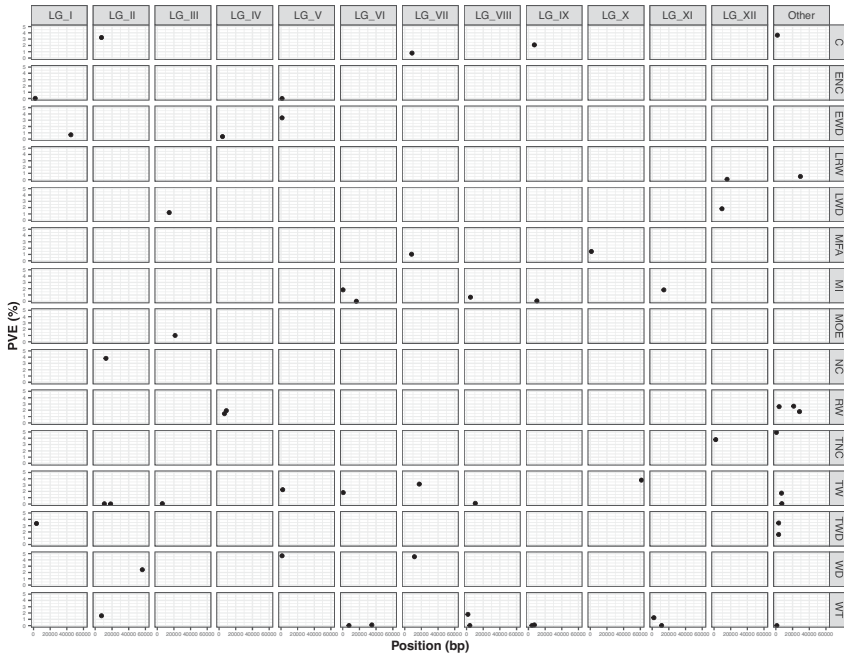
Figure 3. Trait variance explained by QTLs in each linkage group

The highest proportion of the variance explained resulted from the additive action of three QTLs on wood density (11.39%), and five QTLs on ring width (10.4%). For the other traits, the percentage of variance explained by QTLs was lower than 10%. This result clearly indicates that GWAS failed to detect the additional QTLs responsible of the remaining genetic variance. This lack of power can be attributed to several explanations. Firstly, wood properties are complex traits controlled by a large number of QTLs, each with a small effect (Du et al., 2013; McKown et al., 2014a; Porth et al., 2013; Wegrzyn et al., 2010). Secondly, the rapid decay of LD that is the result of the outcrossing nature of Norway spruce and its large effective population size. In our study, LD displayed diverse patterns among different genes. Additionally, the sequence capture method only covered a total of 2331.1 kbp of the exonic sequence, 2470.9 kbp of the intronic sequence, 40.7 kbp of the UTR-like sequence and 9119 exon−intron boundaries (Vidalis et al ., 2018). Therefore, a large portion of the genome was not represented, and this would be compounded by the rapid LD in spruce.

Another important aspect to be considered is population structure, which can influence LD, causing confounding effects and false positives or

spurious associations between traits and molecular marker alleles (Sawitri et al., 2020). Typically, GWAS methods incorporate tools to correct for population structure, which in return are likely to remove a substantial amount of significant gene association signals (Yengo et al., 2018). In paper II and VI samples were derived from a relatively unstructured population. Despite this, we included in the GWAS model the first two components from the principle component analysis (PCA) which explained 5.3% of the molecular marker differentiation.

In light of our results and previous GWAS studies in forest trees, we conclude that a much larger sample size, in the order of tens of thousands of trees, and a denser SNP markers coverage of genic and non-genic regions would be required to detect a significant proportion of the additional markers (Du et al., 2018). Furthermore, the power of GWAS to identify a true association between an SNP and trait is dependent on how strongly the two allelic variants differ in their phenotypic effect (the effect size), and their allele frequency. In Norway spruce, despite the relatively low levels of population structure, it is evident that there are differences in both pattern of LD and allele frequencies between populations, calling for caution when estimating parameters across a wide-range of sampling studies (Larsson et al., 2013).

4.2.2 Disease resistance

In paper III, GWAS based on a population of 466 plus trees returned 11 new potential SNPs for resistance to *H. parviporum*. Individual SNPs explained similar fractions of the phenotypic variation observed, a finding in agreement with interval mapping based QTL study by Lind (Lind et al., 2014). Based on the newly published ultra-dense genetic map (Bernhardsson et al., 2019), five of the SNPs were found to be independent from each other, supported by their localization at different LGs.

The candidate genes linked to SWG were more often expressed in the sapwood than the genes linked to LL. Two candidate genes (i.e., significant SNPs were located within those genes) were associated with the LL trait in bark, MA_53835g0010 and PaLAC5, which are likely to be members of the gene cascade that induced defence against *H. parviporum*. These candidate genes present new insights into the interaction between Norway spruce and *H.parviporum*, such as the putative involvement of the secretory and endosomal trafficking pathways and the laccase PaLAC5 in the control of

lesion extension in the inner bark, or the potential role of mitochondrial protein import and biogenesis in controlling *H.parviporum* spread in the sapwood.

4.3 GWAS to identify genes correlated with fungal communities and pathogens

In paper IV, we combined internal transcribed spacer sequencing of the fungal communities associated with dormant vegetative buds with a GWAS in 478 unrelated Norway spruce trees. The aim was to detect host loci associated with variation in the fungal communities across the population, and to identify loci correlating with the presence of specific, latent, pathogens.

Operational Taxonomic Units (OTUs) with the highest number of reads were reduced using PCA. The first six PCs of the PCA were used as phenotype to conduct GWAS and six Norway spruce SNPs were found associated with the fungal community composition based on the eigenvalues describing the third, fourth and fifth PC. Three additional SNPs associated with colonization of the Norway spruce needle cast pathogen (*Lirula macrospora*), or the cherry spruce rust (*Thekopsora areolata*) in asymptomatic tissues were detected. This study confirmed that dormant vegetative buds are, like needles, colonized by phyllosphere fungi, as the taxonomic identities of many of the most heavily sequenced OTUs showed affinity to previously reported members of the Norway spruce phyllosphere fungal community.

Several of the potential orthologues of the Norway spruce candidate genes associated with PC3 and PC4 indicate that shoot development and morphogenesis are important factors in shaping genotype - specific phyllosphere communities. Our GWAS results suggested that processes in the morphogenesis and flush of the Norway spruce shoot exert a strong influence on the dominant players in the phyllosphere community detected in dormant buds. So, using genome-wide association as a tool, it will be possible to identify many more causative genes and their roles across linked traits, highlighting the potential for systematically uncovering the balanced regulation between these traits and identifying the key hub genes that link them (Feltus, 2014; Lamara et al., 2016). Genomic methods offer the means to screen large numbers of loci associated with environmental cues.

4.4 Genomic prediction in Norway spruce

In paper V, h^2 and Prediction Abilities (PA) based on ABLUP and GBLUP were compared for density, MFA and MOE which were measured using increment cores at age 19 years, and also for Pilodyn at age 22, Velocity and MOE_{ind} at 24 years, on standing trees.

For density and Pilodyn, similar estimates were obtained with ABLUP and GBLUP, thus following the general trend that has been previously reported in other forest trees (Beaulieu et al., 2014b; El-Dien et al., 2015; Resende Jr et al., 2012). This suggests that larger number of SNPs are required to capture historical LD in a species where LD decays rapidly. However, for MFA, a low heritability trait, the PA estimation based on the GBLUP model was substantially higher (0.16), compared to the ABLUP model (0.04). After standardizing with h, the predictive accuracies (PC) of the methods become more similar across traits. In forest trees, pedigree-based PA estimates in conifer species have been reported to be higher or comparable to molecular marker-based models (Chen et al., 2018a; El-Dien et al., 2015; Lenz et al., 2017; Müller et al., 2017; Zapata-Valenzuela et al., 2013), but there are also some studies reporting molecular marker-based PA estimates to be higher (El-Dien et al., 2018; Kainer et al., 2018; Suontama et al., 2019).

The performance of other molecular marker-based statistical models was similar, thus indicating that a large part of the SNPs had relatively equally effect on wood quality traits. Otherwise, methods such as BayesB, which is more efficient when large-effect QTL exist, should have outperformed the other methods. This is also in line with previous results for growth and wood quality traits in other forest tree species (Beaulieu et al., 2014a; Ratcliffe et al., 2015; Thistlethwaite et al., 2017).

In this study we also estimated PAs using partial increment cores with different depths, with the aim of investigating the minimum number of rings (from bark) required to obtain reliable estimates in order to reduce tree injury. For example, for wood density (in Figure 4) our GS model could be trained at tree ages 10 to 12, including only the three to five outermost rings.
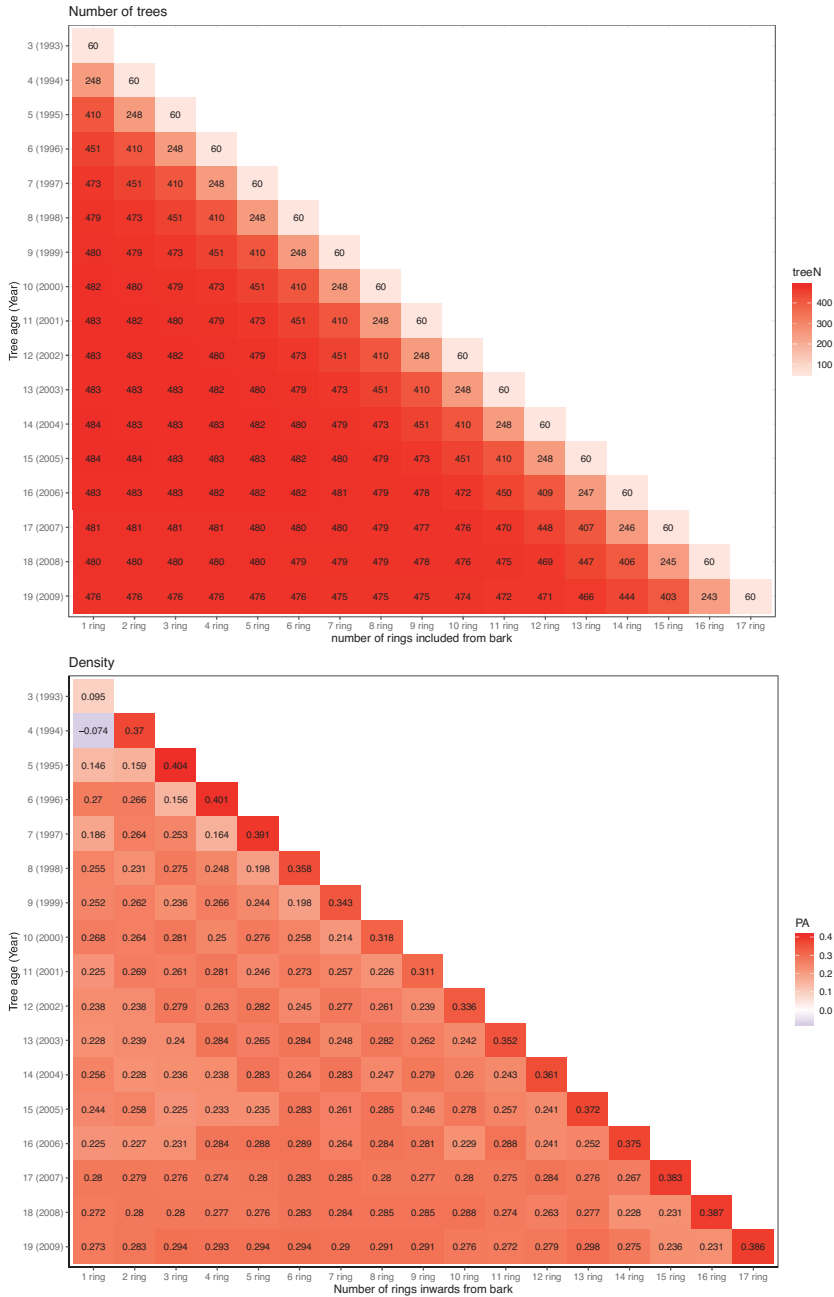
Figure 4. Upper: Number of trees at each tree age with different number of rings. Lower: PA of density from bark to pith at different tree ages (y-axis) and an increasing number of rings included in the estimation (x-axis)

There are four factors that affect accuracy of predictions significantly in the context of GS: (1) The level of LD between molecular markers and QTLs, (2) consanguinity between training and validation population, (3) effective population size (Ne), particularly the number of individuals with the phenotype and genotype used to develop prediction model, and (4) heritability of the trait (Hayes et al., 2009). In our study, molecular markers failed to detect any historical LD beyond the available and often single-generation pedigree information, which may explain why PA was observed to be similar between ABLUP and GBLUP for most of traits. PA estimates generally showed a moderate increase in this study with increasing training set size, irrespective of the statistical method used, since the effective population size increased as more trees were involved in model prediction development. Exceptions to this trend where MFA and MOE estimates were found at 80% of the trees in the training set. In general, the order of PA among different traits was consistent with the order of traits h^2 estimation irrespective of the GS model, which strongly suggests the importance of choosing traits with high heritability as feasible breeding goals, together with experimental designs involving multiple testing sites to build accurate predictive models (Zhang et al., 2017).

5. Conclusions and future perspective

This thesis presents an evaluation of different genomic-based methods to assist Norway spruce breeding. Based on the results of the studies of this thesis work, the following conclusions and perspectives can be drawn:

- Higher repeatability for wood property traits indicates that selection of trees in clonal archives based on phenotypic data represents a cost- and time-effective method.
- Backward selection, whether based solely on offspring data or a combination of offspring and multiple parental clone data, would be most effective for wood density and least effective for MFA.
- GWAS is a valuable genomic-based tool to investigate the genetic architecture of complex traits. However, GWAS full potential would require a substantial scale up in sample size and molecular-marker density in order to become an efficient tool for early marker-assisted selection.
- GS model training for wood quality improvement in Norway spruce should be conducted at a tree age of no younger than 10 to 12 years old.
- GS has the potential to become an alternative for Norway spruce breeding, following a drop in the cost of genotyping. A better reference genome may improve the quality of SNPs so that more information rather than just pedigree could be captured. GS implementation into forward selection would require an acceleration of the flowering time in order to deliver its full potential.
- Fungal community composition in dormant buds of Norway spruce trees is partly determined by the genotype of the host tree.

References

Alqudah, A. M., Sallam, A., Stephen Baenziger, P., Börner, A., 2020. GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley – A review. Journal of Advanced Research 22, 119-135, doi:https://doi.org/10.1016/j.jare.2019.10.013.

Androsiuk, P., Shimono, A., Westin, J., Lindgren, D., Fries, A., Wang, X. R., 2013. Genetic status of Norway spruce (Picea abies) breeding populations for northern Sweden. Silvae Genetica 62, 127-136, doi:https://doi.org/10.1515/sg-2013-0017.

Arnerup, J., Swedjemark, G., Elfstrand, M., Karlsson, B., Stenlid, J., 2010. Variation in growth of Heterobasidion parviporum in a full-sib family of Picea abies. Scandinavian Journal of Forest Research 25, 106-110.

Asiegbu, F. O., Adomas, A., Stenlid, J., 2005. Conifer root and butt rot caused by Heterobasidion annosum (Fr.) Bref. sl. Molecular plant pathology 6, 395-409.

Båga, M., Fowler, D. B., Chibbar, R. N., 2009. Identification of genomic regions determining the phenological development leading to floral transition in wheat (Triticum aestivum L.). Journal of experimental botany 60, 3575-3585.

Baltunis, B. S., Wu, H. X., Powell, M. B., 2007. Inheritance of density, microfibril angle, and modulus of elasticity in juvenile wood of Pinus radiata at two locations in Australia. Canadian Journal of Forest Research 37, 2164-2174.

Bartholomé, J., Van Heerwaarden, J., Isik, F., Boury, C., Vidal, M., Plomion, C., Bouffier, L., 2016. Performance of genomic prediction within and across generations in maritime pine. BMC genomics 17, 604.

Beaulieu, J., Doerksen, T. K., MacKay, J., Rainville, A., Bousquet, J., 2014a. Genomic selection accuracies within and between environments and small breeding groups in white spruce. BMC genomics 15, 1048.

Beaulieu, J., Doerksen, T., Clément, S., MacKay, J., Bousquet, J., 2014b. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. Heredity 113, 343-352.

Beavis, W., Paterson, A., 1998. Molecular dissection of complex traits.

Bernhardsson, C., Vidalis, A., Wang, X., Scofield, D. G., Schiffthaler, B., Baison, J., Street, N. R., García-Gil, M. R., Ingvarsson, P. K., 2019. An Ultra-Dense Haploid Genetic Map for Evaluating the Highly Fragmented Genome Assembly of Norway Spruce (Picea abies). G3: Genes|Genomes|Genetics 9, 1623, doi:10.1534/g3.118.200840.

Bhat, J. A., Ali, S., Salgotra, R. K., Mir, Z. A., Dutta, S., Jadon, V., Tyagi, A., Mushtaq, M., Jain, N., Singh, P. K., Singh, G. P., Prabhu, K. V., 2016.

Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding. Frontiers in genetics 7, 221-221, doi:10.3389/fgene.2016.00221.

Borner, A., Schumann, E., Furste, A., Coster, H., Leithold, B., Roder, S., Weber, E., 2002. Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (Triticum aestivum L.). Theor Appl Genet 105, 921-936, doi:10.1007/s00122-002-0994-1.

Chen, Z.-Q., Gil, M. R. G., Karlsson, B., Lundqvist, S.-O., Olsson, L., Wu, H. X., 2014. Inheritance of growth and solid wood quality traits in a large Norway spruce population tested at two locations in southern Sweden. Tree genetics & genomes 10, 1291-1303.

Chen, Z.-Q., Baison, J., Pan, J., Karlsson, B., Andersson, B., Westin, J., García-Gil, M. R., Wu, H. X., 2018a. Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. BMC Genomics 19, 946, doi:10.1186/s12864-018-5256-y.

Chen, Z.-Q., Lundén, K., Karlsson, B., Vos, I., Olson, Å., Lundqvist, S.-O., Stenlid, J., Wu, H. X., García Gil, M. R., Elfstrand, M., 2018b. Early selection for resistance to Heterobasidion parviporum in Norway spruce is not likely to adversely affect growth and wood quality traits in late-age performance. European Journal of Forest Research 137, 517-525, doi:10.1007/s10342-018-1120-5.

Downes, G. M., Nyakuengama, J. G., Evans, R., Northway, R., Blakemore, P., Dickson, R. L., Lausberg, M., 2002. Relationship between wood density, microfibril angle and stiffness in thinned and fertilized Pinus radiata. Iawa Journal 23, 253-265.

Du, Q., Pan, W., Xu, B., Li, B., Zhang, D., 2013. Polymorphic simple sequence repeat (SSR) loci within cellulose synthase (PtoCesA) genes are associated with growth and wood properties in P opulus tomentosa. New Phytologist 197, 763-776.

Du, Q., Lu, W., Quan, M., Xiao, L., Song, F., Li, P., Zhou, D., Xie, J., Wang, L., Zhang, D., 2018. Genome-Wide Association Studies to Improve Wood Properties: Challenges and Prospects. Frontiers in plant science 9, 1912-1912, doi:10.3389/fpls.2018.01912.

Eckert, A. J., Wegrzyn, J. L., Cumbie, W. P., Goldfarb, B., Huber, D. A., Tolstikov, V., Fiehn, O., Neale, D. B., 2012. Association genetics of the loblolly pine (Pinus taeda, Pinaceae) metabolome. New Phytologist 193, 890-902.

El-Dien, O. G., Ratcliffe, B., Klápště, J., Chen, C., Porth, I., El-Kassaby, Y. A., 2015. Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. BMC genomics 16, 370.

El-Dien, O. G., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., El-Kassaby, Y. A., 2018. Multienvironment genomic variance decomposition analysis of open-pollinated Interior spruce (Picea glauca x engelmannii). Molecular breeding 38, 26.

Falconer, D. S., Mackay, T. F. C., 1996. Introduction to quantitative genetics.

Gao, H., Wu, Y., Li, J., Li, H., Li, J., Yang, R., 2014. Forward LASSO analysis for high-order interactions in genome-wide association study. Briefings in bioinformatics 15, 552-561.

Gion, J.-M., Carouché, A., Deweer, S., Bedon, F., Pichavant, F., Charpentier, J.-P., Baillères, H., Rozenberg, P., Carocha, V., Ognouabi, N., Verhaegen, D., Grima-Pettenati, J., Vigneron, P., Plomion, C., 2011. Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: Eucalyptus. BMC Genomics 12, 301, doi:10.1186/1471-2164-12-301.

Grattapaglia, D., 2017. Status and Perspectives of Genomic Selection in Forest Tree Breeding. In: Varshney, R. K., et al., Eds.), Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement. Springer International Publishing, Cham, pp. 199-249.

Grattapaglia, D., Resende, M. D., 2011. Genomic selection in forest tree breeding. Tree Genetics & Genomes 7, 241-255.

Haines, D. W., Leban, J.-M., 1997. Evaluation of the MOE of Norway spruce by the resonance flexure method. Forest products journal 47, 91.

Hall, D., Hallingbäck, H. R., Wu, H. X., 2016. Estimation of number and size of QTL effects in forest tree traits. Tree Genetics & Genomes 12, 110, doi:10.1007/s11295-016-1073-0.

Hanocq, E., Niarquin, M., Heumez, E., Rousset, M., Le Gouis, J., 2004. Detection and mapping of QTL for earliness components in a bread wheat recombinant inbred lines population. Theoretical and Applied Genetics 110, 106-115.

Hayes, B., Goddard, M., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-1829.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., Goddard, M. E., 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92, 433-43, doi:10.3168/jds.2008-1646.

Hellgren, M. B., Stenlid, J., 1995. Long-term reduction in the diameter growth of butt rot affected Norway spruce, Picea abies. Forest Ecology and Management 74, 239-243, doi:https://doi.org/10.1016/0378-1127(95)03530-N.

Heslot, N., Yang, H.-P., Sorrells, M. E., Jannink, J.-L., 2012. Genomic selection in plant breeding: a comparison of models. Crop science 52, 146-160.

Hiraoka, Y., Fukatsu, E., Mishima, K., Hirao, T., Teshima, K. M., Tamura, M., Tsubomura, M., Iki, T., Kurita, M., Takahashi, M., Watanabe, A., 2018. Potential of Genome-Wide Studies in Unrelated Plus Trees of a Coniferous Species, Cryptomeria japonica (Japanese Cedar). Frontiers in plant science 9, 1322-1322, doi:10.3389/fpls.2018.01322.

Ingvarsson, P. K., Hvidsten, T. R., Street, N. R., 2016. Towards integration of population and comparative genomics in forest trees. New Phytologist 212, 338-344, doi:10.1111/nph.14153.

Isik, F., Bartholomé, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., Plomion, C., Bouffier, L., 2016. Genomic selection in maritime pine. Plant Science 242, 108-119.

Jermstad, K. D., Bassoni, D. L., Wheeler, N. C., Anekonda, T. S., Aitken, S. N., Adams, W. T., Neale, D. B., 2001. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. II. Spring and fall cold-hardiness. Theoretical and Applied Genetics 102, 1152-1158, doi:10.1007/s001220000506.

Kainer, D., Stone, E. A., Padovan, A., Foley, W. J., Külheim, C., 2018. Accuracy of genomic prediction for foliar terpene traits in Eucalyptus polybractea. G3: Genes, Genomes, Genetics 8, 2573-2583.

Karlsson, B., Rosvall, O., 1993. Progeny testing and breeding strategies. In: Lee, S. J., (Ed.), Forestry commission, Edinburgh.

Knowles, R. L., Hansen, L. W., Wedding, A., Downes, G., 2004. Evaluation of non-destructive methods for assessing stiffness of Douglas fir trees. New Zealand Journal of Forestry Science 34, 87-101.

Kumar, N., Kulwal, P., Balyan, H., Gupta, P., 2007. QTL mapping for yield and yield contributing traits in two mapping populations of bread wheat. Molecular Breeding 19, 163-177.

Laidò, G., Marone, D., Russo, M. A., Colecchia, S. A., Mastrangelo, A. M., De Vita, P., Papa, R., 2014. Linkage Disequilibrium and Genome-Wide Association Mapping in Tetraploid Wheat (Triticum turgidum L.). PLOS ONE 9, e95211, doi:10.1371/journal.pone.0095211.

Lamara, M., Raherison, E., Lenz, P., Beaulieu, J., Bousquet, J., MacKay, J., 2016. Genetic architecture of wood properties based on association analysis and co-expression networks in white spruce. The New phytologist 210, 240-255, doi:10.1111/nph.13762.

Larsson, H., Källman, T., Gyllenstrand, N., Lascoux, M., 2013. Distribution of long-range linkage disequilibrium and Tajima's D values in Scandinavian populations of Norway Spruce (Picea abies). G3 (Bethesda, Md.) 3, 795-806, doi:10.1534/g3.112.005462.

Lenz, P., MacKay, J., Rainville, A., Cloutier, A., Beaulieu, J., 2011. The influence of cambial age on breeding for wood properties in Picea glauca. Tree genetics & genomes 7, 641-653.

Lenz, P., Auty, D., Achim, A., Beaulieu, J., Mackay, J., 2013. Genetic improvement of white spruce mechanical wood traits—early screening by means of acoustic velocity. Forests 4, 575-594.

Lenz, P. R., Beaulieu, J., Mansfield, S. D., Clément, S., Desponts, M., Bousquet, J., 2017. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (Picea mariana). BMC genomics 18, 335.

Li, Z., Sillanpää, M. J., 2015. Dynamic quantitative trait locus analysis of plant phenomic data. Trends in plant science 20, 822-833.

Lind, M., Källman, T., Chen, J., Ma, X.-F., Bousquet, J., Morgante, M., Zaina, G., Karlsson, B., Elfstrand, M., Lascoux, M., Stenlid, J., 2014. A Picea abies Linkage Map Based on SNP Markers Identifies QTLs for Four Aspects of Resistance to Heterobasidion parviporum Infection. PLOS ONE 9, e101049, doi:10.1371/journal.pone.0101049.

Lindström, H., Harris, P., Nakada, R., 2002. Methods for measuring stiffness of young trees. Holz als Roh-und Werkstoff 60, 165-174.

Lundqvist, S.-O., Seifert, S., Grahn, T., Olsson, L., García-Gil, M. R., Karlsson, B., Seifert, T., 2018. Age and weather effects on between and within ring variations of number, width and coarseness of tracheids and radial growth of young Norway spruce. European journal of forest research 137, 719-743.

Lynch, M., Walsh, B., 1998. Genetics and analysis of quantitative traits.

Maccaferri, M., Sanguineti, M. C., Corneti, S., Ortega, J. L. A., Salem, M. B., Bort, J., DeAmbrogio, E., del Moral, L. F. G., Demontis, A., El-Ahmed, A., 2008. Quantitative trait loci for grain yield and adaptation of durum wheat (Triticum durum Desf.) across a wide range of water availability. Genetics 178, 489-511.

Markussen, T., Tusch, A., Stephan, B. R., Fladung, M., 2004. Identification of Molecular Markers for Selected Wood Properties of Norway Spruce Picea abies L. (Karst.) I. Wood Density 53, 45, doi:https://doi.org/10.1515/sg-2004-0008.

Marone, D., Russo, M. A., Laidò, G., De Vita, P., Papa, R., Blanco, A., Gadaleta, A., Rubiales, D., Mastrangelo, A. M., 2013. Genetic basis of qualitative and quantitative resistance to powdery mildew in wheat: from consensus regions to candidate genes. BMC genomics 14, 562.

Marone, D., Del Olmo, A. I., Laidò, G., Sillero, J. C., Emeran, A. A., Russo, M. A., Ferragonio, P., Giovanniello, V., Mazzucotelli, E., De Leonardis, A. M., 2009. Genetic analysis of durable resistance against leaf rust in durum wheat. Molecular breeding 24, 25-39.

McCough, S. R., Doerge, R. W., 1995. QTL mapping in rice. Trends in Genetics 11, 482-487, doi:https://doi.org/10.1016/S0168-9525(00)89157-X.

McKown, A. D., Klápště, J., Guy, R. D., Geraldes, A., Porth, I., Hannemann, J., Friedmann, M., Muchero, W., Tuskan, G. A., Ehlting, J., 2014a. Genome‐wide association implicates numerous genes underlying ecological trait variation in natural populations of Populus trichocarpa. New Phytologist 203, 535-553.

McKown, A. D., Klápště, J., Guy, R. D., Geraldes, A., Porth, I., Hannemann, J., Friedmann, M., Muchero, W., Tuskan, G. A., Ehlting, J., Cronk, Q. C. B., El-Kassaby, Y. A., Mansfield, S. D., Douglas, C. J., 2014b. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of Populus trichocarpa. New Phytologist 203, 535-553, doi:10.1111/nph.12815.

Meuwissen, T. H. E., Hayes, B. J., Goddard, M. E., 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157, 1819-1829.

Müller, D., Schopp, P., Melchinger, A. E., 2017. Persistency of prediction accuracy and genetic gain in synthetic populations under recurrent genomic selection. G3: Genes| Genomes| Genetics 7, 801.

Navabi, A., Tewari, J., Singh, R., McCallum, B., Laroche, A., Briggs, K., 2005. Inheritance and QTL analysis of durable resistance to stripe and leaf rusts in an Australian cultivar, Triticum aestivum'Cook'. Genome 48, 97-107.

Neale, D. B., Savolainen, O., 2004. Association genetics of complex traits in conifers. Trends in plant science 9, 325-330.

Neale, D. B., Kremer, A., 2011. Forest tree genomics: growing resources and applications. Nat Rev Genet 12, 111-22, doi:10.1038/nrg2931.

O'Connor, K., Hayes, B., Hardner, C., Nock, C., Baten, A., Alam, M., Henry, R., Topp, B., 2020. Genome-wide association studies for yield component traits in a macadamia breeding population. BMC Genomics 21, 199, doi:10.1186/s12864-020-6575-3.

Olmos, S., Distelfeld, A., Chicaiza, O., Schlatter, A., Fahima, T., Echenique, V., Dubcovsky, J., 2003. Precise mapping of a locus affecting grain protein content in durum wheat. Theoretical and Applied Genetics 107, 1243-1251.

Panio, G., Motzo, R., Mastrangelo, A., Marone, D., Cattivelli, L., Giunta, F., De Vita, P., 2013. Molecular mapping of stomatal‐conductance‐related traits in durum wheat (Triticum turgidum ssp. durum). Annals of applied biology 162, 258-270.

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., Buerkle, C. A., 2012. Genome‐wide association genetics of an adaptive trait in lodgepole pine. Molecular ecology 21, 2991-3005.

Pavy, N., Lamothe, M., Pelgas, B., Gagnon, F., Birol, I., Bohlmann, J., Mackay, J., Isabel, N., Bousquet, J., 2017. A high-resolution reference genetic map positioning 8.8 K genes for the conifer white spruce: structural genomics implications and correspondence with physical distance. The Plant Journal 90, 189-203, doi:10.1111/tpj.13478.

Peng, J., Ronin, Y., Fahima, T., Roder, M. S., Li, Y., Nevo, E., Korol, A., 2003. Domestication quantitative trait loci in Triticum dicoccoides, the progenitor of wheat. Proc Natl Acad Sci U S A 100, 2489-94, doi:10.1073/pnas.252763199.

Plomion, C., Bastien, C., Bogeat-Triboulot, M.-B., Bouffier, L., Déjardin, A., Duplessis, S., Fady, B., Heuertz, M., Le Gac, A.-L., Le Provost, G., Legué, V., Lelu-Walter, M.-A., Leplé, J.-C., Maury, S., Morel, A., Oddou-Muratorio, S., Pilate, G., Sanchez, L., Scotti, I., Scotti-Saintagne, C., Segura, V., Trontin, J.-F., Vacher, C., 2016. Forest tree genomics: 10 achievements from the past 10 years and future prospects. Annals of Forest Science 73, 77-103, doi:10.1007/s13595-015-0488-3.

Porth, I., Klapšte, J., Skyba, O., Hannemann, J., McKown, A. D., Guy, R. D., DiFazio, S. P., Muchero, W., Ranjan, P., Tuskan, G. A., 2013. Genome‐wide association mapping for wood characteristics in P opulus identifies an

array of candidate single nucleotide polymorphisms. New Phytologist 200, 710-726.

Pot, D., Rodrigues, J.-C., Rozenberg, P., Chantre, G., Tibbits, J., Cahalan, C., Pichavant, F., Plomion, C., 2006. QTLs and candidate genes for wood properties in maritime pine (Pinus pinaster Ait.). Tree Genetics & Genomes 2, 10-24, doi:10.1007/s11295-005-0026-9.

Procunier, J. D., Gray, M. A., Howes, N. K., Knox, R. E., Bernier, A. M., 1997. DNA markers linked to a T10 loose smut resistance gene in wheat (Triticum aestivum L.). Genome 40, 176-179, doi:10.1139/g97-025.

Ratcliffe, B., El-Dien, O. G., Klápště, J., Porth, I., Chen, C., Jaquish, B., El-Kassaby, Y., 2015. A comparison of genomic selection models across time in interior spruce (Picea engelmannii× glauca) using unordered SNP imputation methods. Heredity 115, 547-555.

Resende Jr, M., Muñoz, P., Acosta, J., Peter, G., Davis, J., Grattapaglia, D., Resende, M., Kirst, M., 2012. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytologist 193, 617-624.

Resende, M. D., Resende, M. F., Sansaloni, C. P., Petroli, C. D., Missiaggia, A. A., Aguiar, A. M., Abad, J. M., Takahashi, E. K., Rosado, A. M., Faria, D. A., 2012. Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytologist 194, 116-128.

Risch, N., Merikangas, K., 1996. The future of genetic studies of complex human diseases. Science 273, 1516-1517.

Rosvall, O., 2019. Using Norway spruce clones in Swedish forestry: Swedish forest conditions, tree breeding program and experiences with clones in field trials. Scandinavian Journal of Forest Research 34, 342-351, doi:10.1080/02827581.2018.1562566.

Rosvall, O., Bradshaw, R. H. W., Egertsdotter, U., Ingvarsson, P. K., Wu, H., 2019. Using Norway spruce clones in Swedish forestry: introduction. Scandinavian Journal of Forest Research 34, 333-335, doi:10.1080/02827581.2018.1562565.

Ruppert, D., Wand, M. P., Carroll, R. J., 2003. Semiparametric regression. Cambridge university press.

Sawitri, N, T., M, N. i., Widiyatno, S, I., K, U., R, S., KKS, N., SL, L., Y, T., 2020. Potential of Genome-Wide Association Studies and Genomic Selection to Improve Productivity and Quality of Commercial Timber Species in Tropical Rainforest, a Case Study of Shorea platyclados. Forests.

Scallan, A., Green, H., 2007. A technique for determining the transverse dimensions of the fibres in wood. Wood and Fiber Science 5, 323-333.

Skogforsk, 2011. Review of the swedish tree breeding programme. In: Rosvall, O., (Ed.).

Stenlid, J., Swedjemark, G., 1988. Differential growth of S-and P-isolates of Heterobasidion annosum in Picea abies and Pinus sylvestris. Transactions of the british mycological society 90, 209-213.

Sun, X., Marza, F., Ma, H., Carver, B. F., Bai, G., 2010. Mapping quantitative trait loci for quality factors in an inter-class cross of US and Chinese wheat. Theoretical and applied genetics 120, 1041-1051.

Suontama, M., Klápště, J., Telfer, E., Graham, N., Stovold, T., Low, C., McKinley, R., Dungey, H., 2019. Efficiency of genomic prediction across two Eucalyptus nitens seed orchards with different selection histories. Heredity 122, 370-379.

Tan, B., Grattapaglia, D., Martins, G. S., Ferreira, K. Z., Sundberg, B., Ingvarsson, P. K., 2017. Evaluating the accuracy of genomic prediction of growth and wood traits in two Eucalyptus species and their F 1 hybrids. BMC plant biology 17, 110.

Thistlethwaite, F. R., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., Stoehr, M. U., El-Kassaby, Y. A., 2017. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. BMC genomics 18, 930.

Todorovska, E., Christov, N., Slavov, S., Christova, P., Vassilev, D., 2009. Biotic stress resistance in wheat—breeding and genomic selection implications. Biotechnology & Biotechnological Equipment 23, 1417-1426.

VanRaden, P. M., 2008. Efficient methods to compute genomic predictions. Journal of dairy science 91, 4414-4423.

Vidalis, A., Scofield, D. G., Neves, L. G., Bernhardsson, C., García-Gil, M. R., Ingvarsson, P., 2018. Design and evaluation of a large sequence-capture probe set and associated SNPs for diploid and haploid samples of Norway spruce (Picea abies). bioRxiv, 291716.

Wegrzyn, J. L., Eckert, A. J., Choi, M., Lee, J. M., Stanton, B. J., Sykes, R., Davis, M. F., Tsai, C. J., Neale, D. B., 2010. Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (Populus trichocarpa, Salicaceae) secondary xylem. New Phytologist 188, 515-532.

White, T. L., Hodge, G. R., 2013. Predicting breeding values with applications in forest tree improvement. Springer Science & Business Media.

Whittaker, J. C., Thompson, R., Denham, M. C., 2000. Marker-assisted selection using ridge regression. Genetics Research 75, 249-252.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J., Visscher, P. M., Consortium, G., 2018. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Human molecular genetics 27, 3641-3649, doi:10.1093/hmg/ddy271.

Young, N. D., Kumar, L., Menancio-Hautea, D., Danesh, D., Talekar, N. S., Shanmugasundarum, S., Kim, D. H., 1992. RFLP mapping of a major bruchid resistance gene in mungbean (Vigna radiata, L. Wilczek). TAG.

Theoretical and applied genetics. Theoretische und angewandte Genetik 84, 839-844, doi:10.1007/bf00227394.

Zapata-Valenzuela, J., Whetten, R. W., Neale, D., McKeand, S., Isik, F., 2013. Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. G3: Genes, Genomes, Genetics 3, 909-916.

Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., Cui, Z., Ruan, Y., Burgueño, J., San Vicente, F., Olsen, M., Prasanna, B. M., Crossa, J., Yu, H., Zhang, X., 2017. Effect of Trait Heritability, Training Population Size and Marker Density on Genomic Prediction Accuracy Estimation in 22 bi-parental Tropical Maize Populations. Frontiers in Plant Science 8, doi:10.3389/fpls.2017.01916.

Zobel, B. J., van Buijtenen, J. P., 1989. Wood variation: its causes and control. Springer-Verlag, Berlin
.

Popular science summary

Tall parents usually have tall children. This can be explained with genetics, the science that study heredity, and genes are the functional unit of heredity. People's height is determined by a certain number of genes which are inherited by the next generation. As in humans, when crossing two tall trees the result is also tall offspring trees. Tree breeders are well-aware of this phenomenon and is for that reason that they invest a lot of effort in identifying the best trees to be crossed.

In Europe, Norway spruce is one of the most economic important species for being the main source of wood for construction and pulp. This explains the efforts that breeders conduct to maintain spruce forests healthy, while increasing their growth and wood quality. Traditionally breeders are able to succeed in their search for the tallest trees with the only use of two tools, a meter and statistics. Recent technology offers the possibility to select the tallest trees without measuring them. That magic is called Genomic Selection (GS). GS can predict how tall a tree will be just by measuring gene effects with mathematical tools, rather than measuring the tree height with a meter. It is a genomic-based powerful tool that aims to accelerate and improve tree selection in a cost-effective way.

This thesis is pioneer to investigate the feasibility of genomic-based breeding for Norway spruce in Sweden.

Populärvetenskaplig sammanfattning

Långa föräldrar får långa barn. Detta kan förklaras med genetik, den vetenskap som studerar nedärvning, och gener är den funktionella enheten för nedärvning. Människans kroppslängd bestäms av ett visst antal gener som nedärvs till nästa generation. Liksom för människor ger korsningar av två långa träd långa avkommor. Detta är något som skogsträdsförädlare är medvetna om och av detta skäl satsas mycket kraft på att identifiera de bästa träden för samkorsning.

Gran (*Picea abies*) är ett av Europas ekonomiskt viktigaste trädslag som råvara för konstruktionsvirke och pappersmassa. Detta förklarar de ansträngningar som förädlare gör för att bibehålla granens anpassning och vitalitet samtidigt som tillväxt och vedkvalitet förbättras. Det traditionella sättet för förädlare att identifiera träd med den bästa tillväxten är att använda mätstång eller höjdmätare samt statistik. Nyutvecklad teknologi ger dock möjlighet att välja träd med den bästa tillväxten utan att mäta dem. Denna teknologi är genom-baserad förädling. Istället för att mäta träden med någon mätutrustning mäts generna som kontrollerar höjdtillväxten. Med andra ord, genombaserad förädling kan bedöma hur högt ett träd kommer att bli genom att bara studera dess gener. Genombaserat urval är ett kraftfullt redskap som syftar till att snabba på och förbättra selektionen av träd på ett kostnadseffektivt sätt.

Denna avhandling är banbrytande då den studerar möjligheterna för genombaserad förädling av gran i Sverige.

Acknowledgements

My first gratitude belongs to my main supervisor **María Rosario García Gil** for offering me the opportunity to be a PhD student in UPSC, SLU. **Rosario**, thanks for your guidance in research, support in both work and daily life. Whenever I have a request, you just gave me two words "go ahead". I also would like to express my appreciation to my co-supervisors **Harry Wu, Bo Karlson and Mikko J. Sillanpää** for your valuable recommendation and help. I thank all my co-authors, without you I cannot complete my study so smoothly. Especially **Malin Elfstrand** and **Sven-Olof Lundqvist**, your passion and cautious attitude motivating me all the time.

In addition, I would sincerely thank everyone previously or currently in our group. **John Baison**, you opened the gate of bioinformatics to me and was never tired of answering my tons of questions. **Zhiqiang Chen**, you taught me a lot of knowledge of quantitative genetics and also mushrooms collection with limited words. **Haleh Hayatgheibi**, you are such a lovely girl that make my research life more cheerful. **Anders Fries**, thanks for your help for field work, you are so energetic and experienced. **Tomas Funda** and **Irena Fundová**, I think you two are perfect partners and I have to learn the wise for both research and life from you. **Henrik Hallingbäck**, thanks for those valuable discussions we had, I think everyone in the group benefits from your concrete and solid knowledge a lot. **Sonali Ranade**, you look like my warm big sister and saved me several times when I forgot something. **Tianyi Liu, Yuting Zhang** and **Junjie Zhang**, thank you for teaching and helping me with DNA extractions, without that, I may still do extraction now. **Ainhoa Calleja-Rodriguez**, **Hong Nguyen** and **Jenny Lundströmer**, thanks for sharing all information and experience with me.

Many thanks to all colleagues working in UPSC who creating such a warm and friendly environment. Especially our department administrators **Inga-Lis Johansson**, **Lena Gustavsson**, **Anders Eman,** communications

officer **Anne Honsel** and IT support **Simon Birve**. You always solve any of my problems with patience and smile.

Last but not least, I have to express my grateful to my mother who always stand on my side unconditionally. Thanks to my husband who let me be myself and become a better me; two little gentlemen who continuously working very hard to keep my life busy enough. And you, who looking at me as stars in the sky.

I

NRC Research Press

ARTICLE

Genetic analysis of wood quality traits in Norway spruce open-pollinated progenies and their parent plus trees at clonal archives and the evaluation of phenotypic selection of plus trees

Linghua Zhou, Zhiqiang Chen, Sven-Olof Lundqvist, Lars Olsson, Thomas Grahn, Bo Karlsson, Harry X. Wu, and María Rosario García-Gil

Abstract: A two-generation pedigree involving 519 Norway spruce (*Picea abies* (L.) Karst.) plus trees (at clonal archives) and their open-pollinated (OP) progenies was studied with the aim to evaluate the potential of plus-tree selection based on phenotype data scored on the plus trees. Two wood properties (wood density and modulus of elasticity, MOE) and one fiber property (microfibril angle, MFA) were measured with a SilviScan instrument on samples from one ramet per plus tree and 12 OP progenies per plus tree (total of 6288 trees). Three ramets per plus tree and their OP progenies were also assessed for Pilodyn penetration depth and Hitman acoustic velocity, which were used to estimate MOE. The narrow-sense heritability (h^2) estimates based on parent–offspring regression were marginally higher than those based on half-sib correlation when three ramets per plus tree were included. For SilviScan data, estimates of the correlation between half-sib, progeny-based breeding values (BVs) and plus-tree phenotypes, as well as repeatability estimates, were highest for wood density, followed by MOE and MFA. Considering that the repeatability estimates from the clonal archive trees were higher than any h^2 estimate, selection of the best clones from clonal archives would be an effective alternative.

Key words: solid wood, Norway spruce, parent–offspring regression, heritability, repeatability.

Résumé : Une population pedigree de deux générations comprenant 519 arbres plus d'épicéa commun (*Picea abies* (L.) Karst.; d'archives clonales) et leurs descendants issus de pollinisation libre (OP) ont été étudiés conjointement dans le but d'évaluer le potentiel de sélection d'arbres plus en fonction de données phénotypiques prises sur ces derniers. Deux propriétés du bois (densité du bois et module d'élasticité, MOE) et une propriété des fibres (angle des microfibrilles, MFA) ont été mesurées avec un instrument SilviScan sur les échantillons d'un ramet par arbre plus et 12 descendants issus d'OP par arbre plus (total de 6288 arbres). Trois ramets par arbre plus et leur descendants d'OP ont également été évalués pour la profondeur de pénétration du Pilodyn et la vitesse acoustique à l'aide d'un appareil Hitman, afin d'estimer le MOE. Les valeurs d'héritabilité au sens strict (h^2) basées sur la relation parents–progéniture étaient marginalement plus élevées que celles basées sur la corrélation de demi-fratries, lorsque trois ramets par arbre plus étaient considérés. Pour les données de SilviScan, les estimations de la corrélation entre les valeurs en croisement (BV) découlant de l'analyse des demi-fratries et les phénotypes d'arbres plus, ainsi que les estimations de répétabilité, étaient les plus élevées pour la densité de bois, suivie par MOE et MFA. Considérant que les estimations de répétabilité découlant des arbres d'archives clonales étaient plus élevées que toutes les valeurs de h^2, la sélection des meilleurs clones à partir d'archives clonales apparaît comme une alternative efficace. [Traduit par la Rédaction]

Mots-clés : bois massif, épicéa commun, régression parents–progéniture, héritabilité, répétabilité.

Introduction

Norway spruce (*Picea abies* (L.) Karst.) is one of the most important conifer species in Europe for both economic and ecological aspects (Spiecker 2000). Higher volume production, vitality, and log quality for straightness and branch angle have traditionally been the main objectives of the species breeding program, while more recently, different aspects related to wood quality are gaining increasing attention (Mullin et al. 2011; Rosvall et al. 2011). For

mechanical properties of wood-based products, wood density, microfibril angle (MFA), and modulus of elasticity (MOE) are considered as the most important solid-wood quality traits (Chen et al. 2015; Zobel and Jett 1995), and therefore they are the focus of our study.

The SilviScan technology was developed to measure radial variation (i.e., from pith to bark) of solid-wood quality traits, including wood density, MFA, and MOE (Evans 1999, 2008; Evans and Elic

L. Zhou, Z. Chen, and M.R. García-Gil. Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Science, Umeå, Sweden.
S.-O. Lundqvist. IIC, Rosenlundsgatan 48B, SE-118 63 Stockholm, Sweden; RISE Bioeconomy, Box 5604, SE-114 86 Stockholm, Sweden.
L. Olsson and T. Grahn. RISE Bioeconomy, Box 5604, SE-114 86 Stockholm, Sweden.
B. Karlsson. Skogforsk, Ekebo 2250, SE-268 90 Svalöv, Sweden.
H.X. Wu. Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Science, Umeå, Sweden; CSIRO NRCA, Black Mountain Laboratory, Canberra, ACT 2601, Australia.
Corresponding author: María Rosario García-Gil (email: m.rosario.garcia@slu.se).

Published at www.nrcresearchpress.com/cjfr on 6 February 2019.

2001), as well as fiber traits (Evans 1994). Its high efficiency compared with that of corresponding laboratory methods contributed substantially to advances in research and development within wood biology, forestry, and the optimal use of forest resources in softwoods (Lindström et al. 1998; Lundgren 2004; Kostiainen et al. 2009; McLean et al. 2010; Piispanen et al. 2014; Fries et al. 2014), in hardwoods (Kostiainen et al. 2014; Lundqvist et al. 2017), and on modelling of trait variations (Wilhelmsson et al. 2002; Lundqvist et al. 2005; Franceschini et al. 2012; Auty et al. 2014). SilviScan is also used to produce benchmark data and validate the procedures of more rapid and nondestructive methods. Examples of solid-wood traits are Pilodyn penetration depth and Hitman acoustic velocity (hereafter referred to as Pilodyn and velocity, respectively; Chen et al. 2015; Kennedy et al. 2013; Vikram et al. 2011). Pilodyn is an indirect, nondestructive, low-cost, and easy-to-use instrument for estimating wood density. In Norway spruce and other conifer species, strong genetic correlations have been observed between Pilodyn penetration depth and wood density measured with SilviScan (Chen et al. 2015; Cown 1978; Desponts et al. 2017; Fukatsu et al. 2011; King et al. 1988; Sprague et al. 1983; Yanchuk and Kiss 1993). Further, acoustic velocity measured with Hitman apparatus has been shown as an efficient, indirect method related to MFA and has already been used on many species, including Scots pine (*Pinus sylvestris* L.; Hong et al. 2015), white spruce (*Picea glauca* (Moench) Voss; Lenz et al. 2013), and Norway spruce (Chen et al. 2015). Models for many species were implemented in an earlier version of SilviScan (Evans and Elic 2001), followed by further improvements (Evans 2008). An analogous model using the proxy measurements of acoustic velocity and Pilodyn penetration on standing trees was shown to be efficient for selection based on wood stiffness in Norway spruce (Chen et al. 2015). Pilodyn, however, measures wood density in only the outermost annual rings; therefore, it has also been suggested that it may not be reliable for ranking the whole tree in cases where the correlation between the outermost rings and inner rings is low (Wessels et al. 2011) or if the diameter of tree is wide.

A common practice in forest tree breeding programs, which aims to guarantee early genetic gain, is to phenotypically select superior genotypes (plus trees) from naturally regenerated mature stands (Zobel and Talbert 1984; Danusevicius and Lindgren 2002). In Sweden, selection of the breeding base population of Norway spruce plus trees started in the 1940s (Karlsson and Rosvall 1993). Presently, large numbers of plus trees are maintained in ex situ, grafted clonal archives. These archives serve as breeding base populations in which crossings of selected parental genotypes are conducted with the purpose of generating cross-pollinated progenies for the next generation in the breeding cycle. After establishment of the clonal archives, the plus trees are genetically evaluated (ranked) for growth, straightness, branch angle, and vitality superiority following the backward-selection approach. Backward selection is an expensive method that starts with the establishment of open-pollinated (OP) progenies for large numbers of families in progeny trials often tested at multiple sites. This is followed by the assessment of the progenies at more than one site and at a tree age high enough for selection and finally the estimation of breeding values (BVs) to identify the superior genotypes (White et al. 2007). A less expensive alternative to backward selection is the direct selection of plus trees in the clonal archives based on phenotype data directly measured on the plus trees. This approach can be incorporated as the first part of a two-stage selection approach in which plus trees are first selected based on phenotype data for traits of high heritability, followed by a second selection based on clonal or progeny testing (Danusevicius and Lindgren 2005).

The goal of this study is to evaluate the potential of selection based on phenotype data of outstanding plus trees compared with backward selection based on OP progeny trials. For this, we conducted the following three analyses:

1. correlations between the plus-tree BVs for wood density, MFA, and MOE estimated based on OP progenies and plus-tree phenotypes measured at the clonal archive; where SilviScan-based data were available, correlations were estimated for each annual ring;
2. narrow-sense heritability (h^2) based on parent–offspring regression and half-sib progeny correlation; and
3. repeatability or the proportion of clone variation at the clonal archive to conduct plus-tree selection.

Materials and methods

Plant material

The study was based on a two-generation pedigree involving 519 mother plus trees from two different clonal archives located at Ekebo and Maltesholm in southern Sweden. The clonal archive at Ekebo was established in 1984 and the one at Maltesholm was established in 1985–1987. At the time of establishment, 10 ramets on average were grafted for each plus tree and planted with a spacing of 3 m × 0.5 m. At the time of sampling, spacing had been increased through thinning two times, leaving the majority of the genotypes with first seven and then only three ramets remaining. For their corresponding 519 OP families, more progenies per family were planted at each progeny trial. Data from two progeny trials were used: S21F9021146 (also known as F1146; Höreda, Eksjö, Sweden) and S21F9021147 (also known as F1147; Erikstorp, Tollarp, Sweden), both established in 1990 with a spacing of 1.4 m × 1.4 m. The same OP families were present in both progeny trials. Increment cores from the progenies of the OP families were sampled in 2010 and from the ramets at the clonal archive in 2017.

Silvicultural activities

Mild precommercial thinnings were conducted in Höreda and Erikstorp in 2008, at the age of 18 years, and in 2010, at the age of 20 years. At the first thinning, only strongly suppressed trees that were judged to not reach commercial dimensions were cut down. Most of these were less than 50 mm diameter at breast height (DBH; breast height = 1.30 m), and their removal was assumed to have no effect on the growth or properties of the remaining trees. The second thinning was performed in the year of sampling and influenced only the outermost growth ring, for which data were excluded for other reasons (see following sections). The clone archives at Ekebo and Maltesholm were topped in autumn 2007 at age 23 years, when a large seed crop was harvested. The uppermost 15%–20% of the trees was removed. Thinnings of the Ekebo clonal archive and parts of the Maltesholm archive were carried out for the first time in the late 1990s and the last time in autumn 2009 at age 25 years.

Phenotypic measurements

The radial variations in wood density, MFA, and MOE had been assessed already in a previous study (Chen et al. 2014). Increment cores of up to 12 progenies per OP family (six from each progeny trial) had been analyzed from pith to bark with SilviScan, followed by the calculation of area-weighted means, representing the trait means of all wood formed in the stem cross sections at each cambial age. In the current study, analogous SilviScan data were also generated for one ramet from each clone from the parental 519 plus trees at the clonal archives. Pilodyn 6J Forest and Hitman ST300 instruments were used on the standing trees to assess penetration depth and acoustic velocity (respectively) of the same ramets. These measurements were used to estimate MOE as the indirect methods (MOE(ind)) using the following formula:

$$MOE(ind) = (1/Pilo) \times 10000 \times AV^2$$

where Pilo is the Pilodyn penetration depth (mm), and AV is the velocity of the wave through the material (km·s⁻¹). AV has a strong

Fig. 1. Mean values generated with SilviScan data from the open-pollinated (OP) progenies and the clonal archive. MFA, microfibril angle; MOE, modulus of elasticity.
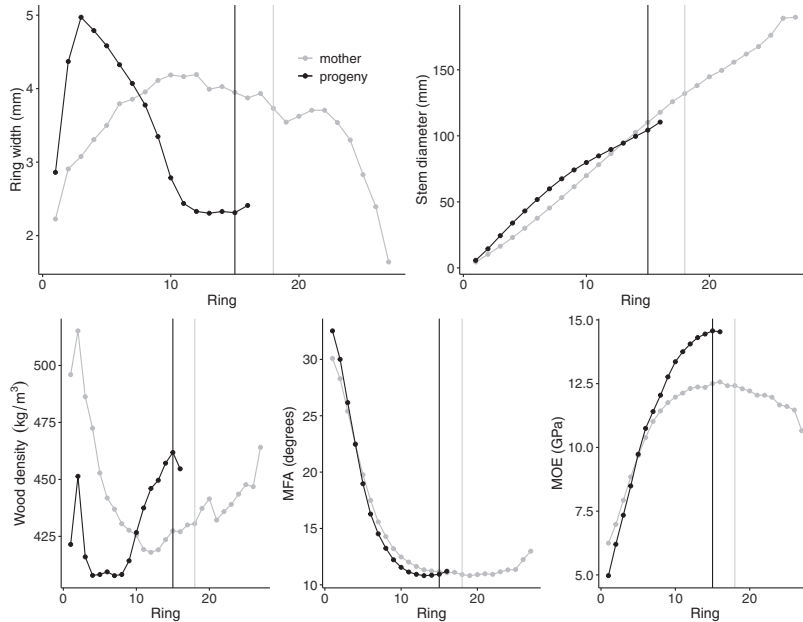
inverse correlation with MFA, and the inverse of Pilo has a strong correlation with wood density (Chen et al. 2015).

When data for more than one ramet were available, the mean was used for further Pearson correlation analysis. The evaluations were based on data from ring 3 to ring 16. The two rings closest to the pith were removed from the evaluations, as the rings here may be so curved that the X-ray beam used on measurement will pass through wood of adjacent rings. However, values for rings 1 and 2 are kept in Fig. 1 to illustrate the described problem. Data on rings larger than 16 from the progeny trials were excluded to avoid problems of representability, given that the slow-growing trees did not reach the highest cambial ages (Lundqvist et al. 2018). The number of rings per tree varied from 10 to 18. Further, data for the outermost ring of each tree were excluded from the evaluations, as they may not be fully formed, to avoid problems of data distortion due to damage of the ring during the increment core extraction.

The genetic parameters were calculated based on means for stem cross sections at different cambial ages (ring numbers) using R (version 3.3.3; R Core Team 2017).

BV of mothers based on progeny tests

The linear mixed model used for the estimation of parental BV and variance components was expressed in matrix form:

$$y = Xb + Zu + e$$

where y is a vector of measured data, b is a vector of fixed effects with design matrix X, u is a vector of random effects with design matrix Z, and e is a vector of residuals. Fixed- and random-effect solutions were obtained by solving the following mixed-model equation (White and Hodge 2013):

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I\alpha \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

where b is the fixed effects, including site, block within site, and provenance; u is the random effect, which is the family; I is the identity matrix with dimensions equal to the number of mothers; and α is a ratio of residual variance and genetic variance explained by the random family effect.

The estimations of BV (u), variance, and covariance components were done using the lme4 package (Bates et al. 2015) in R (version 3.3.3; R Core Team 2017).

Pearson correlation

For all wood properties measured with SilviScan and indirect methods, Pearson correlation was calculated between the plus trees' BVs and plus trees' phenotype data. In the case of SilviScan-based analysis, only one ramet was available, whereas in the case of Pilodyn, velocity, and MOE(ind), two or three ramets were available, depending on the OP family.

Narrow-sense heritability (h^2)

Two methods for calculating heritability were estimated. The first method was based on half-sib family progeny analysis, and the linear mixed model was fitted as follows:

$$y_{ijklm} = \mu + S_i + B_{j(i)} + P_k + F_{l(k)} + SF_{il(k)} + e_{ijklm}$$

where y_{ijklm} is the phenotypic individual observation; μ is the general mean; S_i, $B_{j(i)}$, and P_k are the fixed effects of site i, block j, and provenance k, respectively; $F_{l(k)}$ is the random effect of family l within provenance k; $SF_{il(k)}$ is the random interactive effect of site i

and family l within provenance k; and e_{ijklm} is the random residual effect for individual tree m.

Narrow-sense heritability was estimated for each trait as

$$\hat{h}^2 = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_P^2} = \frac{4 \times \hat{\sigma}_F^2}{\hat{\sigma}_F^2 + \hat{\sigma}_{SF}^2 + \hat{\sigma}_e^2}$$

where $\hat{\sigma}_A^2$, $\hat{\sigma}_P^2$, $\hat{\sigma}_F^2$, $\hat{\sigma}_{SF}^2$, and $\hat{\sigma}_e^2$ are estimations of additive genetic variance (A), phenotypic variance (P), family variance (F), family–site interaction variance (SF), and residual variance (e), respectively.

The second method was based on parent–offspring regression. We used a linear regression to model the mother–offspring pairs for each trait value:

$$Y = \beta_0 + \beta_1 X$$

where Y is the phenotype value for the offspring, β_0 is the intercept of the regression, β_1 is the slope of the regression, and X is the phenotype value for a mother. Because genetic covariance between parents and offspring is equal to $(1/2)\sigma_A^2$ (Falconer and Mackay 1996), we can get

$$\beta_1 = \frac{\text{cov}(X, Y)}{\text{var}\,X} = \frac{(1/2)\sigma_A^2}{\sigma_P^2}$$

The individual tree h^2 is

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}$$

So from the slope of the regression, the estimation of h^2 can be obtained from

$$\hat{h}^2 = 2\hat{\beta_1}$$

The standard error of heritability is estimated by $2/\sqrt{N}$ (Falconer and Mackay 1996), where N is the number of families. This way, the parent–offspring-based heritability was computed for SilviScan data for each annual ring and for Pilodyn, velocity, and MOE(ind). To allow comparison between the estimates based on SilviScan and those based on indirect measurements, all heritabilities were computed only on the 162 families for which three ramets were available in the clonal archives.

In our study, the heritabilities for SilviScan data were calculated for each cambial age from the area-weighted means representing stem cross sections.

Repeatability

Repeatability indicates the proportion of total variation in a trait that is due to differences among clones (Falconer and Mackay 1996). The individual repeatability R was calculated as (Falconer and Mackay 1996; Lynch and Walsh 1998)

$$R = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_e^2}$$

where σ_c^2 is the estimated clone variance, and σ_e^2 is the residual variance.

Progeny size effect on heritability

To investigate the effect of progeny size on the estimation of heritability based on a parent–offspring regression, we used a subset of progeny trees in which each family had exactly six progenies in each of the two trials. In total, 180 families and 2160 progeny trees were included in the analysis. From this subset, one to six progenies were randomly selected per family from each site. The process was bootstrapped 500 times, and the means and standard errors of heritability were then estimated for comparison. The most prominent consequence of increasing the number of OP progenies was the decrease in the standard errors (i.e., more precise estimation of heritability) (Fig. 2). When a progeny size of four trees was selected, parent–offspring heritability stabilized for MOE(ind) and peaked for velocity, whereas it reached a maximum value for Pilodyn at progeny size six. Based on these results, all of the genetic parameters involving progeny data were estimated using the highest number of progeny size.

Results

Traits curve for progenies and plus trees

Mean values for ring width, DBH, wood density, MOE, and MFA were plotted against each annual ring for progenies and plus trees (Fig. 1). Ring numbers larger than 27 for the clonal archive and ring numbers larger than 16 for progeny trees were excluded, as they were based on very few trees.

Ring width and wood density curves showed clear discrepancies between the trees at the clonal archive and those at the progeny trials. In the progeny test, the mean widths of the rings decreased steeply until about ring 10, after which it became rather stable until the overrepresentation of fast-growing trees became visible at above ring 15 (Lundqvist et al. 2018), which is indicated in Fig. 1 with a black, vertical line. The density mean was high closest to the pith, then stable at a low level until ring 10, after which it started to increase steeply until the fast-growing trees became overrepresented. In contrast to the progeny trial, the ring width means of the clonal archives started low and increased steadily until rings 10–12, presumably at the time when the archive was first thinned from dense to low spacious compared with the progeny trials. Then, the means started to decrease with age. These trees were topped at age 23 years, which should approximately correspond to ring 18, indicated with a grey, vertical line. At higher ages, ring width experienced a sharp drop, which can be interpreted as a physiological response of the trees to the removal of the upper canopy. From this, we concluded that data at higher ages of the clonal archive may not represent the natural development of trees and may not be fully comparable with the expected response in the progeny trials at older ages. At ages deemed representative, the wood density curve for the clonal archive mirrored the changes in ring width, which is not surprising considering that growth and density are negatively correlated (Chen et al. 2014). In reference to DBH, we observed that the trees at the clonal archive were thinner from pith up to ring 14. After this ring, they became thicker than those at the progeny trial because of steadily wider rings.

The curves representing change in MFA with annual ring were very similar between the trees at the progeny trial and those at the clonal archive. In both types of plantation, MFA decreased sharply and stabilized towards the bark. The slight increases of the means for the last rings shown may reflect overrepresentation of fast-growing trees. As expected, the decrease in MFA is accompanied by an increase in MOE because of the strong negative correlation between the traits, also shown based on the same data by Chen et al. (2014). It was also expected that the progeny trial MOE would reach higher values than those at the clonal archive, as MOE shows positive correlation with wood density, which is higher for these trees in rings larger than 10. In contrast to ring width and

Fig. 2. Heritability estimation by parent–offspring regression based on different numbers of progenies for Pilodyn penetration depth, Hitman acoustic velocity, and MOE(ind). The number of ramets per mother clone varied among plus trees from one to three. [Colour online.]
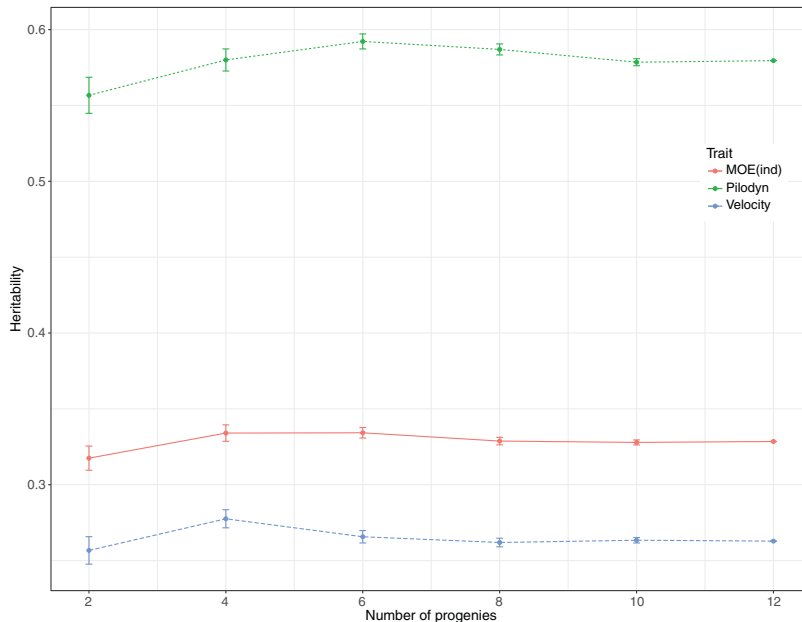
wood density, MFA and MOE curves did not reveal an effect of tree topping.

BV and phenotypic value correlation

Per-ring correlations between half-sib, progeny-based BVs and plus-tree phenotypes for the SilviScan data are presented in Fig. 3. For wood density, correlation estimates increased steadily from low levels at the pith to about 0.4 at rings 12–15. For MFA, the estimates reached a plateau of about 0.17 at rings 4–7 and then decreased gradually. The estimates for MOE were in-between: an initial increase was followed by a plateau, with a decreasing tendency only near the bark, which was possibly an effect of the increasing overrepresentation of fast-growing trees at those ring numbers.

The estimated correlations between half-sib, progeny-based BVs and plus-tree phenotypes were 0.29, 0.13, and 0.23 for Pilodyn, velocity, and MOE(ind), respectively. When using three plus-tree ramets, the correlations increased to 0.32, 0.15, and 0.28, respectively. These values were in concordance with the SilviScan-based estimates of correlation, in which the highest values were reached for density, followed by MOE and MFA.

Heritability estimates on progeny and parent–offspring regression

Estimations of h^2 based on parent–offspring regression at each annual ring using SilviScan data are presented in Fig. 4. The h^2 estimates for wood density increased from pith to bark, and for MFA, they remained on the same level across all annual rings. For MOE, an initial increase of the h^2 estimates was followed by a plateau.

The h^2 estimations of the whole-stem cross sections based on half-sib progeny correlation and parent–offspring regression are presented in Table 1. Based on progeny correlation, the h^2 estimates were 0.43, 0.29, and 0.38 for wood density, MFA, and MOE,

respectively. For mean parent–offspring, the h^2 estimates (based on one ramet) were 0.35, 0.15, and 0.28 for wood density, MFA, and MOE, respectively. The h^2 values estimated by progeny correlation were 0.31, 0.20, and 0.28 for Pilodyn, velocity, and MOE(ind), respectively. Moreover, based on parent–offspring regression, the h^2 values ranged from 0.27 to 0.41, 0.13 to 0.29, and 0.13 to 0.30 for Pilodyn, velocity, and MOE(ind), respectively. With respect to the indirect measurements of wood quality, these results indicate that h^2 estimations based on parent–offspring regression were only marginally higher than those based on half-sib correlation, even when three ramets per plus tree were included in the analyses. Based on data collected with indirect methods, the progeny-based h^2 estimates were higher than parent–offspring regression h^2 estimates for one ramet. Instead, the progeny-based h^2 estimates were marginally lower than the h^2 estimates obtained for parent–offspring regression for three ramets. Based on SilviScan data, the progeny-based h^2 estimates were higher than the h^2 estimates obtained for parent–offspring regression for one ramet. To allow comparison, all of the h^2 estimates in Table 1 were computed only on the 162 families for which three ramets were available in the clonal archive. Repeatability estimates were higher than any h^2 estimate.

Discussion

In this study, we evaluated the potential of ranking and selection for better solid-wood quality traits of outstanding phenotypes (plus trees) as an alternative to backward selection based on BV estimates on half-sib progenies. The evaluation was based on multiple genetic parameters: correlation between half-sib progeny BVs and plus-tree phenotype data, repeatability, and narrow-sense heritability (h^2) based on parent–offspring regression as compared with half-sib correlation.

Fig. 3. Correlations of SilviScan data for each annual ring between breeding values (BVs) of plus trees estimated from the progeny and area-weighted phenotypic values from the plus trees. [Colour online.]
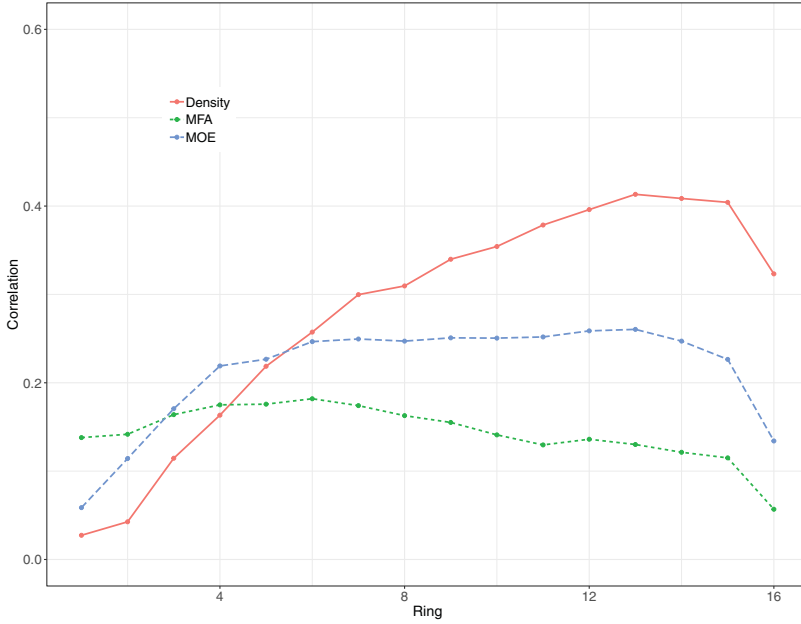
Fig. 4. Heritability estimates using parent–offspring regression of area-weighted values calculated from SilviScan data for each annual ring. [Colour online.]

Published by NRC Research Press
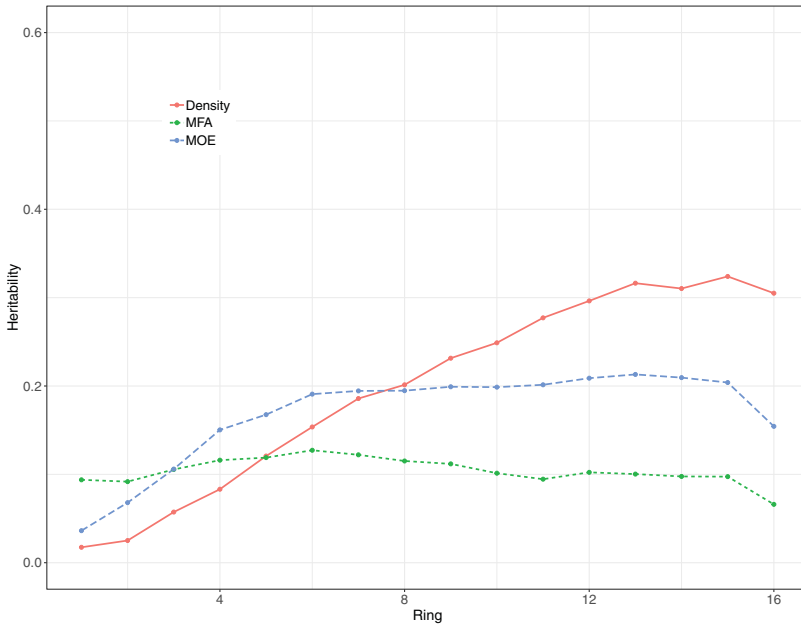
Table 1. Heritability and repeatability estimates based on measurements of wood density, MFA, and MOE from SilviScan, as well as Pilodyn penetration depth and Hitman acoustic velocity, which were used to estimate MOE as the indirect methods.

Methods	SilviScan			Indirect methods		
	Density	MFA	MOE	Pilodyn	Velocity	MOE(ind)
Parent–offspring regression (1 ramet)	**0.35** (±0.16)	**0.15** (±0.16)	**0.28** (±0.16)	0.27 (±0.15)	0.13 (±0.15)	0.13 (±0.15)
Parent–offspring regression (3 ramets)	N/A	N/A	N/A	**0.41** (±0.15)	**0.29** (±0.15)	**0.30** (±0.15)
Half-sib correlation (offspring only)	**0.43** (±0.09)	**0.29** (±0.08)	**0.38** (±0.08)	**0.31** (±0.08)	**0.20** (±0.08)	**0.28** (±0.08)
Repeatability	N/A	N/A	N/A	**0.52** (±0.06)	**0.30** (±0.04)	**0.45** (±0.05)

Note: To allow comparison, all heritability estimates (± standard error) were based only on the 162 families for which three ramets were available in the clonal archive. Estimates that are statistically significantly different from zero are indicated in boldface type. MFA, microfibril angle; MOE, modulus of elasticity; MOE(ind), modulus of elasticity for indirect methods; N/A, not applicable.

The h^2 estimates for wood density, MFA, and MOE measured with SilviScan from increment cores were 0.43, 0.29, and 0.38 and 0.35, 0.15, and 0.28 based on progeny correlation and parent–offspring regression, respectively. When using indirect measurements directly on standing trees, the h^2 estimates based on progeny correlation were 0.31, 0.20, and 0.28 for Pilodyn, velocity, and MOE(ind), respectively. Moreover, based on parent–offspring regression, the values ranged from 0.27 to 0.41, 0.13 to 0.29, and 0.13 to 0.30 for Pilodyn, velocity, and MOE(ind), respectively. Our h^2 values estimated by progeny correlation were in the range of those previously reported for wood properties in loblolly pine (*Pinus taeda* L.; Isik et al. 2011), maritime pine (*Pinus pinaster* Aiton; Louzada 2003; Gaspar et al. 2008), lodgepole pine (*Pinus contorta* Douglas ex Loudon; Hayatgheibi et al. 2017), Norway spruce (Hylen 1997, 1999; Hannrup et al. 2004; Hallingbäck et al. 2008), white spruce (Lenz et al. 2010), and British Columbia's interior spruce (Ivkovich et al. 2002). Similarly, our h^2 estimates based on parent–offspring regression also agree with previously reported values for wood properties in Norway spruce (Steffenrem et al. 2016), loblolly pine (Loo et al. 1984; Williams and Megraw 1994), and Sakhalin spruce (*Picea glehnii* (F. Schmidt) Mast.; Tanabe et al. 2015).

Our repeatability estimates for the indirect measurements based on the analysis of three ramets per plus tree were 0.52, 0.30, and 0.45 for Pilodyn, velocity, and MOE(ind), respectively. Previously reported repeatability estimates for wood quality and growth in Norway spruce (Rosner et al. 2007; Gräns et al. 2009; Steffenrem et al. 2016) and Sakhalin spruce (Tanabe et al. 2015) are also in concordance with our estimates, whereas other studies have reported either higher MFA and MOE values in radiata pine (*Pinus radiata* D. Don; Lindström et al. 1998) or lower MOE values in Sitka spruce (*Picea sitchensis* (Bong.) Carrière; Hansen and Roulund 1997).

Interpretation of the discrepancies between progeny and plus tree for ring width and wood properties

The observed discrepancies in developments across annual rings between the trees at the progeny trials and those at the clonal archive for ring width, wood density, and MOE could be attributed to a difference in spacing, including thinning of the clonal archive. During the first years, the trees of the clonal archive were only 0.5 m apart and were under strong competition compared with the trees in the progeny trials. This is presumed to explain their thinner annual rings and higher wood density at these ages. The thinning performed at two occasions even out this difference in competition, and widths and densities become similar. Thinning results in more favourable growth conditions for the clonal archive trees regarding access to light and other resources, which is presumed to explain why trees at these ages instead have wider annual rings and lower wood densities. After topping of the trees, it is harder to relate this to the developments of growth and patterns.

Less spacing among trees is known to result in stronger competition for resources. Under tight spacing, lower diameter is primarily the result of competition for light (Turner et al. 2009). Trees tend to grow taller at the expense of diameter in their at-

tempt to outcompete the neighbouring trees in search of light. Multiple studies in conifer species have reported effects of plantation density on growth (diameter and slenderness) and wood and fiber properties. Wider spacing at planting has been reported to be associated with higher tree diameter and lower MOE in Scots pine (Persson et al. 1995) and a number of coniferous species (Chuang and Wang 2001; Zhang et al. 2002; Clark et al. 2008; Lasserre et al. 2008, 2009; Schimleck et al. 2018). Ring width and wood density are negatively correlated, and MOE is negatively correlated with both wood density and MFA (Loo et al. 1984; Hodge and Purnell 1993; Zhang and Morgenstern 1995; Waghorn et al. 2007; Gaspar et al. 2008; Lasserre et al. 2009; Chen et al. 2014). The effect of spacing on growth and wood properties, together with their well-documented correlations, strengthens our previous interpretation regarding thinner rings and higher density for the clonal archive trees in the first rings and the reverse later on. It also supports our interpretation of the higher MOE, and lower MFA, at these latter ages for the progeny trees.

Although narrow spacing could account for the results that we obtained, it is also possible that additional factors have contributed to the discrepancies between the two plantation types: abiotic factors such as rainfall, temperature, or soil properties. However, a previous study conducted on the same data from the progeny trials, both treated with similar silvicultural activities, revealed low genotype–environment interaction (Chen et al. 2014), which indicates that climatic conditions or soil properties are not factors behind the differences (at least in southern Sweden, where all three plantations are located).

Potential for selection of Norway spruce plus trees on phenotype data at clonal archives

In operational breeding, selection of plus trees as gene donors to the next generation is usually conducted through evaluation of their OP progenies grown in common-garden experiments (progeny trials), a breeding design known as a backward selection. This is a method that involves multiple actions such as seedling production, seedling establishment (often in multiple sites), and assessment and evaluation of multiple tree properties when the trees in the trial have grown at least 10 rings at breast height. The high demands in time and costs of this approach motivate evaluation of alternatives such as plus-tree selection based on phenotype data assessed at the clonal archive.

Phenotypic selection of plus trees is a common practice for establishing the foundations of a breeding program, while providing early genetic gains (Zobel and Talbert 1984). Furthermore, two-stage selection strategies of plus trees, in which plus trees are first selected based on phenotype followed by a second stage based on clonal or progeny test, have previously been proposed in conifers (Danusevicius and Lindgren 2002, 2004). Danusevicius and Lindgren concluded that when heritability is high, phenotypic selection is a superior breeding strategy and a two-stage strategy based on progeny testing improves by the first stage of phenotypic selection.

Considering that repeatability and h^2 estimates are similar, we suggest that selection of MFA at the clonal archive would be an

effective alternative. However, given the low values of correlation among plantations, h^2, and repeatability, a lower efficiency in tree improvement is expected for MFA than for other traits with higher h^2 (e.g., density; Chen et al. 2014). This conclusion can be extended to selection based on both progeny and plus-tree phenotype. The heritability of MOE using three ramets based on parent–offspring regression (0.30) is higher than using half-correlation (0.28); however, considering that clonal repeatability for MOE (0.45) is higher than any h^2 estimate, we suggest that it would be more cost- and time-effective to select clonal archive trees based on MOE scored with indirect measurements. Previously, MFA and MOE have been reported to have low and moderate heritabilities, respectively, in Norway spruce (Hannrup et al. 2004; Lenz et al. 2010; Chen et al. 2014), whereas higher heritabilities have been reported for MOE in Scots pine (Hong et al. 2015) and for MOE and MFA in loblolly pine (Isik et al. 2011). Similar to the other wood properties (repeatability for wood density is higher than correlation and h^2), selection of trees at the clonal archive based on indirect measurements of this trait will be efficient. Considering that h^2 increases towards the bark, a higher response to selection is expected at older ages. Other studies also support our observation of higher heritability for wood density than for MFA and MOE (Lenz et al. 2010; Isik et al. 2011; Chen et al. 2014).

Conclusion

Our study resulted in the following conclusions.

- Narrow spacing at the clonal archive could account for the discrepancies between the progeny trial and clonal archive for ring width and wood density traits.
- Narrow-sense heritabilities (h^2) estimated from parent–offspring regression using a single ramet were lower than those based on half-sib correlation. Based on indirect measurements, parent–offspring h^2 estimates using three ramets were higher than those based on half-sib correlation, indicating that multiple copies of ramets are critical in estimating reliable genetic parameters and making selection in archive.
- Wood density, or its surrogate trait Pilodyn measurement, had the highest h^2 among the three wood quality traits, whether it was based on SilviScan data using increment cores or indirect measurements on standing trees and parent–offspring regression or half-sib correlation, followed by MOE and MFA.
- Backward selection, whether based on offspring data alone or a combination of offspring and clonal archive data, would be most effective for wood density and least effective for MFA.
- Based on higher repeatability estimates as compared with the h^2 estimates, selection of the best clones from clonal archives would be highly cost- and time-effective.
- The observed discrepancies between both plantation types for growth, wood, and fiber properties could be mostly explained by the tighter tree spacing at the clonal archive.

Acknowledgements

References

Auty, D., Achim, A., Macdonald, E., Cameron, A.D., and Gardiner, B.A. 2014. Models for predicting wood density variation in Scots pine. Forestry, **87**(3): 449–458. doi:10.1093/forestry/cpu005.

Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. **67**(1). doi:10.18637/jss.v067.i01.

Chen, Z.-Q., García-Gil, M.R., Karlsson, B., Lundqvist, S.-O., Olsson, L., and Wu, H. 2014. Inheritance of growth and solid wood quality traits in a large Norway spruce population tested at two locations in southern Sweden. Tree Genet. Genomes, **10**(5): 1291–1303. doi:10.1007/s11295-014-0761-x.

Chen, Z.-Q., Karlsson, B., Lundqvist, S.-O., García-Gil, M.R., Olsson, L., and Wu, H. 2015. Estimating solid wood properties using Pilodyn and acoustic velocity on standing trees of Norway spruce. Ann. For. Sci. **72**(4): 499–508. doi:10.1007/s13595-015-0458-9.

Chuang, S.-T., and Wang, S.-Y. 2001. Evaluation of standing tree quality of Japanese cedar grown with different spacing using stress-wave and ultrasonic-wave methods. J. Wood Sci. **47**(4): 245–253. doi:10.1007/BF00766709.

Clark, A., III, Jordan, L., Schimleck, L., and Daniels, R.F. 2008. Effect of initial planting spacing on wood properties of unthinned loblolly pine at age 21. For. Prod. J. **58**(10): 78–83.

Cown, D. 1978. Comparison of the Pilodyn and torsiometer methods for the rapid assessment of wood density in living trees. N. Z. J. For. Sci. **8**(3): 384–391.

Danusevicius, D., and Lindgren, D. 2002. Efficiency of selection based on phenotype, clone and progeny testing in long-term breeding. Silvae Genet. **51**: 19–26.

Danusevicius, D., and Lindgren, D. 2004. Progeny testing preceded by phenotypic pre-selection — timing considerations. Silvae Genet. **53**(1–6): 20–26. doi:10.1515/sg-2004-0004.

Danusevicius, D., and Lindgren, D. 2005. Optimization of breeding population size for long-term breeding. Scand. J. For. Res. **20**(1): 18–25. doi:10.1080/02827580410019517.

Desponts, M., Perron, M., and DeBlois, J. 2017. Rapid assessment of wood traits for large-scale breeding selection in *Picea mariana* [Mill.] B.S.P. Ann. For. Sci. **74**: 53. doi:10.1007/s13595-017-0646-x.

Evans, R. 1994. Rapid measurement of the transverse dimensions of tracheids in radial wood sections from *Pinus radiata*. Holzforschung, **48**(2): 168–172. doi:10.1515/hfsg.1994.48.2.168.

Evans, R. 1999. A variance approach to the X-ray diffractometric estimation of microfibril angle in wood. Appita J. **52**: 283–289.

Evans, R. 2008. Wood stiffness by X-ray diffractometry. *In* Characterization of the cellulosic cell wall. *Edited by* D.D. Stokke and L.H. Groom. Wiley-Blackwell. pp. 138–146.

Evans, R., and Elic, J. 2001. Rapid prediction of wood stiffness from microfibril angle and density. For. Prod. J. **51**(3): 53–57.

Falconer, D., and Mackay, T. 1996. Introduction to quantitative genetics. 4th ed. Longman.

Franceschini, T., Lundqvist, S.-O., Bontemps, J.-D., Grahn, T., Olsson, L., Evans, R., and Leban, J.-M. 2012. Empirical models for radial and tangential fibre width in tree rings of Norway spruce in north-western Europe. Holzforschung, **66**(2): 219–230. doi:10.1515/HF.2011.150.

Fries, A., Ulvcrona, T., Wu, H.X., and Kroon, J. 2014. Stem damage of lodgepole pine clonal cuttings in relation to wood and fiber traits, acoustic velocity, and spiral grain. Scand. J. For. Res. **29**(8): 764–776. doi:10.1080/02827581.2014.978886.

Fukatsu, E., Tamura, A., Takahashi, M., Fukuda, Y., Nakada, R., Kubota, M., and Kurinobu, S. 2011. Efficiency of the indirect selection and the evaluation of the genotype by environment interaction using Pilodyn for the improvement of wood density in *Cryptomeria japonica*. J. For. Res. **16**(2): 128–135. doi:10.1007/s10310-010-0217-6.

Gaspar, M.J., Louzada, J.L., Silva, M.E., Aguiar, A., and Almeida, M.H. 2008. Age trends in genetic parameters of wood density components in 46 half-sibling families of *Pinus pinaster*. Can. J. For. Res. **38**(6): 1470–1477. doi:10.1139/X08-013.

Gräns, D., Hannrup, B., Isik, F., Lundqvist, S.-O., and McKeand, S. 2009. Genetic variation and relationships to growth traits for microfibril angle, wood density and modulus of elasticity in a *Picea abies* clonal trial in southern Sweden. Scand. J. For. Res. **24**(6): 494–503. doi:10.1080/02827580903280061.

Hallingbäck, H.R., Jansson, G., and Hannrup, B. 2008. Genetic parameters for grain angle in 28-year-old Norway spruce progeny trials and their parent seed orchard. Ann. For. Sci. **65**(3): 301. doi:10.1051/forest:2008005.

Hannrup, B., Cahalan, C., Chantre, G., Grabner, M., Karlsson, B., Bayon, I.L., Jones, G.L., Müller, U., Pereira, H., Rodrigues, J.C., Rosner, S., Rozenberg, P., Wilhelmsson, L., and Wimmer, R. 2004. Genetic parameters of growth and wood quality traits in *Picea abies*. Scand. J. For. Res. **19**(1): 14–29. doi:10.1080/02827580310019536.

Hansen, J.K., and Roulund, H. 1997. Genetic parameters for spiral grain, stem form, Pilodyn and growth in 13 years old clones of Sitka spruce (*Picea sitchensis* (Bong.) Carr.). Silvae Genet. **46**(2-3): 107–113.

Hayatgheibi, H., Fries, A., Kroon, J., and Wu, H.X. 2017. Genetic analysis of lodgepole pine (*Pinus contorta*) solid wood quality traits. Can. J. For. Res. **47**(10): 1303–1313. doi:10.1139/cjfr-2017-0152.

Hodge, G.R., and Purnell, R.C. 1993. Genetic parameter estimates for wood density, transition age, and radial growth in slash pine. Can. J. For. Res. **23**(9): 1881–1891. doi:10.1139/x93-238.

Hong, Z., Fries, A., Lundqvist, S.-O., Gull, B.A., and Wu, H.X. 2015. Measuring stiffness using acoustic tool for Scots pine breeding selection. Scand. J. For. Res. **30**(4): 363–372. doi:10.1080/02827581.2014.1001783.

Hylen, G. 1997. Genetic variation of wood density and its relationship with growth traits in young Norway spruce. Silvae Genet. 46(1): 55.

Hylen, J. 1999. Age trends in genetic parameters of wood density in young Norway spruce. Can. J. For. Res. 29(1): 135–143. doi:10.1139/x98-170.

Isik, F., Whetten, R., Zapata-Valenzuela, J., Ogut, F., and McKeand, S. 2011. Genomic selection in loblolly pine — from lab to field. BMC Proc. 5(Suppl. 7): I8. doi:10.1186/1753-6561-5-S7-I8.

Ivkovich, M., Namkoong, G., and Koshy, M. 2002. Genetic variation in wood properties of interior spruce. II. Tracheid characteristics. Can. J. For. Res. 32(12): 2128–2139. doi:10.1139/x02-139.

Karlsson, B., and Rosvall, O. 1993. Breeding programmes in Sweden: Norway spruce. In Progeny Testing and Breeding Strategies: Proceedings of the Meeting of the Nordic Group of Tree Breeding, Edinburgh, UK, 6–10 October 1993. Edited by S.J. Lee. Forestry Commission, Edinburgh, UK. pp. 6–21.

Kennedy, S.G., Cameron, A.D., and Lee, S.J. 2013. Genetic relationships in wood quality traits and diameter growth of juvenile core wood in Sitka spruce. Can. J. For. Res. 43(1): 1–6. doi:10.1139/cjfr-2012-0308.

King, J., Yeh, F., Heaman, J., and Dancik, B. 1988. Selection of wood density and diameter in controlled crosses of coastal Douglas-fir. Silvae Genet. 37(3–4): 152–157.

Kostiainen, K., Kaakinen, S., Saranpää, P., Sigurdsson, B.D., Lundqvist, S.-O., Linder, S., and Vapaavuori, E. 2009. Stem wood properties of mature Norway spruce after 3 years of continuous exposure to elevated [CO_2] and temperature. Global Change Biol. 15(2): 368–379. doi:10.1111/j.1365-2486.2008.01755.x.

Kostiainen, K., Saranpää, P., Lundqvist, S.-O., Kubiske, M.E., and Vapaavuori, E. 2014. Wood properties of Populus and Betula in long-term exposure to elevated CO_2 and O_3. Plant, Cell Environ. 37(6): 1452–1463. doi:10.1111/pce.12261. PMID: 24372544.

Lasserre, J.-P., Mason, E.G., and Watt, M.S. 2008. Influence of the main and interactive effects of site, stand stocking and clone on Pinus radiata D. Don corewood modulus of elasticity. For. Ecol. Manage. 255(8–9): 3455–3459. doi: 10.1016/j.foreco.2008.02.022.

Lasserre, J.-P., Mason, E.G., Watt, M.S., and Moore, J.R. 2009. Influence of initial planting spacing and genotype on microfibril angle, wood density, fibre properties and modulus of elasticity in Pinus radiata D. Don corewood. For. Ecol. Manage. 258(9): 1924–1931. doi:10.1016/j.foreco.2009.07.028.

Lenz, P., Cloutier, A., Mackay, J., and Beaulieu, J. 2010. Genetic control of wood properties in Picea glauca — an analysis of trends with cambial age. Can. J. For. Res. 40(4): 703–715. doi:10.1139/X10-014.

Lenz, P., Auty, D., Achim, A., Beaulieu, J., and Mackay, J. 2013. Genetic improvement of white spruce mechanical wood traits — early screening by means of acoustic velocity. Forests, 4(3): 575–594. doi:10.3390/f4030575.

Lindström, H., Evans, J.W., and Verrill, S.P. 1998. Influence of cambial age and growth conditions on microfibril angle in young Norway spruce (Picea albes [L.] Karst.). Holzforschung, 52(6): 573–581. doi:10.1515/hfsg.1998.52.6.573.

Loo, J.A., Tauer, C.G., and van Buijtenen, J.P. 1984. Juvenile–mature relationships and heritability estimates of several traits in loblolly pine (Pinus taeda). Can. J. For. Res. 14(6): 822–825. doi:10.1139/x84-145.

Louzada, J.L.P.C. 2003. Genetic correlations between wood density components in Pinus pinaster Ait. Ann. For. Sci. 60(3): 285–294. doi:10.1051/forest:2003020.

Lundgren, C. 2004. Cell wall thickness and tangential and radial cell diameter of fertilized and irrigated Norway spruce. Silva Fenn. 38(1): 438. doi:10.14214/sf.438.

Lundqvist, S.-O., Grahn, T., and Hedenberg, Ö. 2005. Models for fibre dimensions in different softwood species. Simulation and comparison of within and between tree variations for Norway and Sitka spruce, Scots and loblolly pine. In Proceedings of the IUFRO 5th Workshop: Wood Quality Modelling, Auckland, New Zealand, 22–27 November 2005. STFI-Packforsk Rep. ART 05/54.

Lundqvist, S.-O., Grahn, T., Olsson, L., and Seifert, T. 2017. Comparison of wood, fibre and vessel properties of drought-tolerant eucalypts in South Africa. South. For. 79(3): 215–225. doi:10.2989/20702620.2016.1254910.

Lundqvist, S.-O., Seifert, S., Grahn, T., Olsson, L., García-Gil, M.R., Karlsson, B., and Seifert, T. 2018. Age and weather effects on between and within ring variations of number, width and coarseness of tracheids and radial growth of young Norway spruce. Eur. J. For. Res. 137(5): 719–743. doi:10.1007/s10342-018-1136-x.

Lynch, M., and Walsh, B. 1998. Genetics and analysis of quantitative traits. Oxford University Press.

McLean, J.P., Evans, R., and Moore, J.R. 2010. Predicting the longitudinal modulus of elasticity of Sitka spruce from cellulose orientation and abundance. Holzforschung, 64(4): 495–500. doi:10.1515/hf.2010.084.

Mullin, T., Andersson, B., Bastien, J.-C., Beaulieu, J., Burdon, R., and Dvorak, W. 2011. Economic importance, breeding objectives and achievements. CRC Press and Edenbridge Science Publishers, New York.

Persson, B., Persson, A., Ståhl, E.G., and Karlmats, U. 1995. Wood quality of Pinus sylvestris progenies at various spacings. For. Ecol. Manage. 76(1–3): 127–138. doi:10.1016/0378-1127(95)03557-Q.

Piispanen, R., Heinonen, J., Valkonen, S., Mäkinen, H., Lundqvist, S.-O., and Saranpää, P. 2014. Wood density of Norway spruce in uneven-aged stands. Can. J. For. Res. 44(2): 136–144. doi:10.1139/cjfr-2013-0201.

R Core Team. 2017. R: a language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org.

Rosner, S., Klein, A., Müller, U., and Karlsson, B. 2007. Hydraulic and mechanical properties of young Norway spruce clones related to growth and wood structure. Tree Physiol. 27(8): 1165–1178. doi:10.1093/treephys/27.8.1165. PMID: 17472942.

Rosvall, O., Stahl, P., Almqvist, C., Anderson, B., Berlin, M., Ericsson, T., Eriksson, M., Gregorsson, B., Hajek, J., and Hallander, J. 2011. Review of the Swedish Tree Breeding Programme. Tech. Rep. Arbetsrapport, Skogforsk, Uppsala Science Park, Uppsala, Sweden.

Schimleck, L., Antony, F., Dahlen, J., and Moore, J. 2018. Wood and fiber quality of plantation-grown conifers: a summary of research with an emphasis on loblolly and radiata pine. Forests, 9(6): 298. doi:10.3390/f9060298.

Spiecker, H. 2000. The growth of Norway spruce (Picea abies [L.] Karst.) in Europe within and beyond its natural range. In Forest Ecosystem Restoration: Ecological and Economical Impacts of Restoration Processes in Secondary Coniferous Forests: Proceedings of the International Conference, Vienna, Austria, 10–12 April 2000. Institute of Forest Growth Research. pp. 247–256.

Sprague, J., Talbert, J., Jett, J., and Bryant, R. 1983. Utility of the Pilodyn in selection for mature wood specific gravity in loblolly pine. For. Sci. 29: 696–701.

Steffenrem, A., Solheim, H., and Skrøppa, T. 2016. Genetic parameters for wood quality traits and resistance to the pathogens Heterobasidion parviporum and Endoconidiophora polonica in a Norway spruce breeding population. Eur. J. For. Res. 135(5): 815–825. doi:10.1007/s10342-016-0975-6.

Tanabe, J., Tamura, A., Ishiguri, F., Takashima, Y., Iizuka, K., and Yokota, S. 2015. Inheritance of basic density and microfibril angle and their variations among full-sib families and their parental clones in Picea glehnii. Holzforschung, 69(5): 581–586. doi:10.1515/hf-2014-0052.

Turner, G.D., Lewis, J.D., Mates-Muchin, J.T., Schuster, W.F., and Watt, L. 2009. Light availability and soil source influence ectomycorrhizal fungal communities on oak seedlings grown in oak- and hemlock-associated soils. Can. J. For. Res. 39(7): 1247–1258. doi:10.1139/X09-051.

Vikram, V., Cherry, M.L., Briggs, D., Cress, D.W., Evans, R., and Howe, G.T. 2011. Stiffness of Douglas-fir lumber: effects of wood properties and genetics. Can. J. For. Res. 41(6): 1160–1173. doi:10.1139/x11-039.

Waghorn, M.J., Mason, E.G., and Watt, M.S. 2007. Influence of initial stand density and genotype on longitudinal variation in modulus of elasticity for 17-year-old Pinus radiata. For. Ecol. Manage. 252(1–3): 67–72. doi:10.1016/j.foreco.2007.06.019.

Wessels, C.B., Malan, F.S., and Rypstra, T. 2011. A review of measurement methods used on standing trees for the prediction of some mechanical properties of timber. Eur. J. For. Res. 130(6): 881–893. doi:10.1007/s10342-011-0484-6.

White, T.L., and Hodge, G.R. 2013. Predicting breeding values with applications in forest tree improvement. Springer Science & Business Media.

White, T.L., Adams, W.T., and Neale, D.B. 2007. Forest genetics. In Forest genetics. CABI. pp. 1–14.

Wilhelmsson, L., Arlinger, J., Spångberg, K., Lundqvist, S.-O., Grahn, T., Hedenberg, Ö., and Olsson, L. 2002. Models for predicting wood properties in stems of Picea abies and Pinus sylvestris in Sweden. Scand. J. For. Res. 17(4): 330–350. doi:10.1080/02827580260138080.

Williams, C.G., and Megraw, R.A. 1994. Juvenile–mature relationships for wood density in Pinus taeda. Can. J. For. Res. 24(4): 714–722. doi:10.1139/x94-095.

Yanchuk, A., and Kiss, G. 1993. Genetic variation in growth and wood specific gravity and its utility in the improvement of interior spruce in British Columbia. Silvae Genet. 42: 141–148.

Zhang, S.Y., and Morgenstern, E.K. 1995. Genetic variation and inheritance of wood density in black spruce (Picea mariana) and its relationship with growth: implications for tree breeding. Wood Sci. Technol. 30(1): 63–75. doi:10.1007/BF00195269.

Zhang, S.Y., Chauret, G., Ren, H.Q., and Desjardins, R. 2002. Impact of initial spacing on plantation black spruce lumber grade yield, bending properties, and MSR yield. Wood Fiber Sci. 34: 460–475.

Zobel, B.J., and Jett, J.B. 1995. Genetics of wood production. Springer-Verlag, Berlin, Heidelberg. doi:10.1007/978-3-642-79514-5.

Zobel, B., and Talbert, J. 1984. Applied forest tree improvement. John Wiley & Sons, New York.

II

Genome-wide association study identified novel candidate loci affecting wood formation in Norway spruce

John Baison[1],* [iD], Amaryllis Vidalis[2], Linghua Zhou[1], Zhi-Qiang Chen[1], Zitong Li[3], Mikko J. Sillanpää[4], Carolina Bernhardsson[1,5], Douglas Scofield[6], Nils Forsberg[1], Thomas Grahn[7], Lars Olsson[7], Bo Karlsson[8], Harry Wu[1], Pär K. Ingvarsson[5,9], Sven-Olof Lundqvist[7,10], Totte Niittylä[1,†] and M Rosario García-Gil[1,*,†]

[1]*Department of Forest Genetics and Plant Physiology, Umeå Plant Science Centre, Swedish University of Agricultural Science, Parallellvägen 21, Umeå 907 36, Sweden,*
[2]*Section of Population Epigenetics and Epigenomics, Centre of Life and Food Sciences Weihenstephan, Technische Universität München, Lichtenbergstr. 2a, München 85748, Germany,*
[3]*Ecological Genetics Research Unit, Department of Biosciences, University of Helsinki, P.O. Box 65, FI-00014, Helsinki, Finland,*
[4]*Department of Mathematical Sciences, Biocenter Oulu, University of Oulu, Pentti Kaiteran katu 1, Oulu, Finland,*
[5]*Department of Ecology and Environmental Science, Umeå University, Linnaeus väg 4-6, Umeå 907 36, Sweden,*
[6]*Uppsala Multidisciplinary Centre for Advanced Computational Science, Uppsala University, Lägerhyddsvägen 2, Uppsala 752 37, Sweden,*
[7]*RISE Bioeconomy, Drottning Kristinas väg 61, SE-114 86, Stockholm, Sweden,*
[8]*Skogforsk, Ekebo 2250, SE-268 90, Svalöv, Sweden,*
[9]*Department of Ecology and Genetics: Evolutionary Biology, Uppsala University, Kåbovägen 4, Uppsala 752 36, Sweden, and*
[10]*IIC, Rosenlundsgatan 48B, SE-118 63, Stockholm, Sweden*

SUMMARY

Norway spruce is a boreal forest tree species of significant ecological and economic importance. Hence there is a strong imperative to dissect the genetics underlying important wood quality traits in the species. We performed a functional genome-wide association study (GWAS) of 17 wood traits in Norway spruce using 178 101 single nucleotide polymorphisms (SNPs) generated from exome genotyping of 517 mother trees. The wood traits were defined using functional modelling of wood properties across annual growth rings. We applied a Least Absolute Shrinkage and Selection Operator (LASSO-based) association mapping method using a functional multilocus mapping approach that utilizes latent traits, with a stability selection probability method as the hypothesis testing approach to determine a significant quantitative trait locus. The analysis provided 52 significant SNPs from 39 candidate genes, including genes previously implicated in wood formation and tree growth in spruce and other species. Our study represents a multilocus GWAS for complex wood traits in Norway spruce. The results advance our understanding of the genetics influencing wood traits and identifies candidate genes for future functional studies.

Keywords: candidate genes, functional trait mapping, genome-wide association mapping, Norway spruce, sequence capture, single nucleotide polymorphisms.

INTRODUCTION

Norway spruce (*Picea abies* (L.) Karst.) is a dominant boreal species of significant economic and ecological importance (Hannrup *et al.*, 2004). Long-term Norway spruce breeding programmes for improvement of growth and survival were initiated in the 1940s and recently, wood quality

has become one of the priority traits (Bertaud and Holmbom, 2004; Hannrup *et al.*, 2004). Norway spruce breeding in Sweden completes one cycle in about 20 years and such long generation time makes improvements in growth and wood quality very slow. Among wood quality traits, wood

83

density is considered a key indicator of stability, strength and stiffness of sawn timber (Hauksson *et al.*, 2001). Several studies of wood quality observed that fast growth conflicts with high quality wood, as shown by the negative genetic correlation between wood volume growth and density in Norway spruce (Olesen, 1977; Dutilleul *et al.*, 1998; Chen *et al.*, 2014). However in several conifers such as Scots pine (*P. sylvestris* L.) and red pine the relationship has been inconsistent (Larocque and Marshall, 1995; Peltola *et al.*, 2009). To combine fast growth and desirable wood properties through breeding, and to shorten the breeding cycle, it is therefore imperative to design effective early selection methods and breeding strategies. In an effort to design optimal breeding and selection strategies it is essential to identify alleles that are responsible for generating favourable or unfavourable genetic correlations (Hallingbäck *et al.*, 2014).

One of the early studies in conifers identified quantitative trait locus (QTLs) for wood density variation in loblolly pine using linkage analyses based on segregating family pedigrees (Groover *et al.*, 1994). However, marker-aided selection (MAS) based on results from QTL linkage analyses were never implemented in practical tree breeding due to the so-called Beavis effect (e.g. inflated estimates of allelic effects and underestimation of QTL number for economically important traits) (Beavis, 1998), inconsistent associations among different families and the low transferability of markers (Strauss *et al.*, 1992). Association mapping (AM), also called linkage disequilibrium (LD) mapping, is a powerful alternative QTL detection method that was introduced to tree genetics using a candidate gene approach (Thumma *et al.*, 2010). AM overcomes the limited resolution of family-based QTL linkage mapping by relying on historical recombination in the mapping population (Neale and Savolainen, 2004; Thavamanikumar *et al.*, 2013; Huang and Han, 2014). However, the genome-wide levels of LD in Norway spruce has been revealed to be complex and highly heterogenous (Larsson *et al.*, 2013). Therefore, AM is also vulnerable to some confounding historical factors such as population admixture, selection pressures which include possible genetic drift. Therefore, population genetic structure, kinship and LD within the study population need to be carefully accounted for in the analysis to minimize false positives (Khan and Korban, 2012).

The availability of a draft genome sequence for Norway spruce (Nystedt *et al.*, 2013) has provided numerous possibilities for the development of genetic markers to conduct both AM at the genome-wide level (genome-wide association, GWAS) and genomic selection (GS). Several reduced representation-based approaches such as sequence capture and transcriptome sequencing (Hirsch *et al.*, 2014) have been developed for studying large genomes, such as the 20 Gb Norway spruce genome. These approaches reduce the sequence space by decreasing the repetitive sequence content of the genome.

Several AM studies have been performed on traits in different tree species and have identified genetic loci linked to, for instance, wood properties in *Populus trichocarpa* Torr. & A. Gray ex. Hook (Porth *et al.*, 2013) and *Eucalyptus* (Resende *et al.*, 2017b), and adaptive traits in *Pinus contorta* Douglas (Parchman *et al.*, 2012). Some genes may impact the trait development at a particular developmental stage, whereas others may alter, or control, rates of change and transitions between consecutive stages (Xing *et al.*, 2012; Anderegg, 2015). Studies aimed at dissecting the genetic basis of such dynamics in wood properties can benefit from the application of mathematical functions that account for year-to-year variation across annual growth rings, cambial age and distance from pith (Li *et al.*, 2014). The development of mathematical methods for the analysis of these longitudinal traits has made it possible to map QTLs underlying the dynamics of developmental traits (Yang *et al.*, 2006; Li and Sillanpää, 2013; Camargo *et al.*, 2018), and to enhance our understanding of the genetic architecture of the growth trajectories of such dynamic traits (Ma *et al.*, 2002; Xing *et al.*, 2012). Such functional mapping analysis can be conducted using a multistage approach (Heuven and Janss, 2010). First, the phenotype trends of each individual are modelled using curve-fitting methods and the parameters describing the curve are then considered as latent traits. The latent traits are then used in an independent association analyses to search for genomic regions affecting the trait and to estimate genetic marker effects (Li and Sillanpää, 2013; Li *et al.*, 2014).

In this study, we applied a functional AM approach to identify genomic regions contributing to wood quality traits in Norway spruce. We applied spline models since traditional analyses that utilise a single point data across annual growth rings may confound the analyses by averaging across a full sample. Such averaging may obscure mechanisms acting at specific time points during wood formation and will make identification of underlying genes more difficult. This study has performed the analysis of number of cells per ring calculated from SilviScan data. Penalized LASSO regression (Tibshirani, 1996) and the stability selection probability method (Meinshausen and Bühlmann, 2010) were then used, to detect significant associations between latent traits derived from estimated breeding values (EBVs) and 178101 SNP markers covering the Norway spruce genome.

RESULTS AND DISCUSSION

All 517 Norway spruce maternal trees in the study were considered for variant detection and an average of 1.5 million paired-end reads were sequenced per individual resulting in 178 101 SNPs. Most SNPs were missense (61%). Applying the probability of stability selection

(SSP) to the intercept, slope and two nodes (β_2 and β_3) we detected 52 significant QTLs in 17 individual traits whose phenotypic variance explained (PVE) ranged from 0.01 to 4.93% (Table 1). 14 of the significant markers were consistent with overdominance ($|d/a| > 1.25$), with the remaining being codominant (27) ($|d/a| < 0.50$) and 10 exhibiting partial to full dominance ($0.50 < |d/a| < 1.25$) (Table 2, Figure 3).

Table 1 Phenotypes, latent traits, SNP, SNP feature, frequency and PVE

Phenotype	Latent Trait	QTL	SNP[a]	Allele	SNP Feature	Frequency	PVE (%)
WD	Intercept	167610	MA_10435406_13733	A/G	Downstream variant	0.71	4.64
	Slope	30469	MA_33109_11804	A/G	Upstream variant	0.72	4.50
	β_2	30469	MA_33109_11804	A/G	Upstream variant	0.551	4.15
	β_3	157442	MA_10432646_63090	G/A	Upstream variant	0.567	2.43
EWD	Intercept	167610	MA_10435406_13733	A/G	Downstream variant	0.545	3.38
	Slope	23798	MA_20321_44812	C/T	Upstream variant	0.53	0.69
		70955	MA_118446_4316	T/A	Upstream variant	0.644	0.40
TWD	Slope	131698	MA_10235390_3386	G/A	Stop gained	0.672	1.58
		160208	MA_10433411_3386	T/C	Intron variant	0.595	3.41
	β_2	89044	MA_212523_6278	T/C	Upstream variant	0.534	3.34
LWD	Slope	43797	MA_62987_13474	T/C	Missense variant	0.524	1.81
		165481	MA_10434805_21408	C/T	Intron variant	0.588	1.21
		171233	MA_10436058_4902	G/A	Intron variant	0.712	4.03
RW	Intercept	11535	MA_10694_9101	A/C	Synonymous variant	0.545	1.95
		112391	MA_879270_7373	C/T,A	Stop gained	0.532	1.45
		112394	MA_879384_3894	C/A	Splice region variant	0.692	2.56
	Slope	165481	MA_10434805_21408	C/T	Intron variant	0.521	2.66
	β_2	23808	MA_20322_28351	T/G	Synonymous variant	0.554	1.78
		165481	MA_10434805_21408	C/T	Intron variant	0.533	0.18
	β_3	23808	MA_20322_28351	T/G	Synonymous variant	0.55	1.20
		165481	MA_10434805_21408	C/T	Intron variant	0.615	1.79
TRW	Slope	111057	MA_817099_1105	T/A	Missense variant	0.685	1.12
	β_2	33110	MA_38472_13803	T/A	Upstream gene variant	0.657	3.23
		89295	MA_214776_1624	G/A	Upstream gene variant	0.688	4.51
	β_3	111057	MA_817099_1105	T/A	Missense variant	0.672	1.20
LRW	Intercept	143628	MA_10428744_29330	C/T	Downstream variant	0.668	0.5
	β_3	164772	MA_10434624_20686	C/A	Downstream variant	0.571	0.06
MOE	Slope	165481	MA_10434805_21408	C/T	Intron variant	0.602	1.00
NC	β_2	145839	MA_10429444_12692	G/C	Upstream variant	0.645	3.82
ENC	Slope	98508	MA_402880_2045	A/C	Upstream variant	0.667	0.03
		167610	MA_10435406_13733	A/G	Downstream variant	0.685	0.01
TNC	Intercept	95870	MA_346723_2241	T/C	Upstream variant	0.667	3.78
		126785	MA_9447489_687	A/C	Upstream gene variant	0.68	4.93
LNC	Intercept	143628	MA_10428744_29330	C/T	Downstream variant	0.66	3.14
	Slope	143628	MA_10428744_29330	C/T	Downstream variant	0.672	4.77
EP	Intercept	16868	MA_15729_40331	G/T	Intron variant	0.609	3.32
		91242	MA_246125_1213	G/A	Synonymous variant	0.594	3.41
TP	Intercept	101203	MA_462319_4322	A/C	Upstream gene variant	0.594	1.16
		132014	MA_10251995_2442	A/C	Upstream gene variant	0.601	3.22
LP	β_2	162397	MA_10434007_77578	C/T	Upstream gene variant	0.892	1.14
EP/LP	Intercept	51657	MA_80954_29644	G/A	Downstream variant	0.63	0.81
		60787	MA_98424_947	C/T	Intron variant	0.655	1.80
		123639	MA_8790100_1384	A/C	Upstream variant	0.628	0.75
	β_2	59480	MA_96191_7122	A/G	Synonymous	0.6	2.37
	β_3	117333	MA_1045136_4310	T/C	Missense variant	0.523	1.34
Mass index (growth × density)	Intercept	166235	MA_10435002_4986	G/A	Intergenic variant	0.533	0.65
	Slope	61096	MA_99004_17108	G/A	Synonymous variant	0.66	0.01
		67181	MA_109804_10278	G/A	Missense variant	0.612	0.05
		1401	MA_1378_4718	C/A	Exon/stop gained	0.588	1.19
		138744	MA_10427214_13968	G/T	Missense variant	0.58	1.80
		162397	MA_10434007_77578	C/T	Upstream variant	0.627	1.44
	β_2	21924	MA_19222_1789	A/G	Upstream variant	0.71	1.82

[a]SNP: The SNP name was composed of the contig (MA_number) and SNP position on contig. For example, the first SNP MA_1043540_13733 was located on contig MA_1043540 at position 13 733 bp; PVE is the phenotypic variance explained.

Table 2 SNP modes of inheritance

Phenotype	QTL	SNP	Allele	2a[a]	d[b]	d/a
WD	167610	MA_10435406_13733	A/G	19.63	9.45	0.96
	30469	MA_33109_11804	A/G	5.19	6.22	2.40
	157442	MA_10432646_63090	G/A	4.49	2.80	1.25
EWD	167610	MA_10435406_13733	A/G	5.495	2.270	0.83
	23798	MA_20321_44812	C/T	4.814	5.905	2.45
	70955	MA_118446_4316	T/A	1.989	0.966	0.97
TWD	131698	MA_10235390_3386	G/A	3.849	−0.636	−0.33
	160208	MA_10433411_3386	T/C	2.313	3.908	3.38
	89044	MA_212523_6278	T/C	0.703	−0.962	−2.73
LWD	43797	MA_62987_13474	T/C	5.124	−1.086	−0.42
	165481	MA_10434805_21408	C/T	4.684	1.165	0.50
	171223	MA_10436058_4902	G/A	0.938	2.482	5.29
RW	11535	MA_10694_9101	A/C	0.111	0.049	0.88
	112391	MA_879270_7373	C/T,A	0.056	−0.027	−0.98
	112394	MA_879384_3894	C/A	0.194	−0.045	−0.45
	165481	MA_10434805_21408	C/T	0.158	0.039	0.49
	23808	MA_20322_28351	T/G	0.025	0.030	2.62
TRW	111057	MA_817099_1105	T/A	0.016	0.001	0.16
	33110	MA_38472_13803	T/A	0.029	−0.002	−0.19
	89295	MA_214776_1624	G/A	0.026	−0.001	−0.13
LRW	143628	MA_10428744_29330	C/T	0.006	−0.002	−0.67
	164772	MA_10434624_20686	C/A	0.002	0.003	2.90
MOE	165481	MA_10434805_21408	C/T	0.376	0.101	0.53
NC	145839	MA_10429444_12692	G/C	0.298	0.792	5.31
ENC	98508	MA_402880_2045	A/C	4.144	1.314	0.63
	167610	MA_10435406_13733	A/G	4.695	−3.033	−1.29
TNC	95870	MA_346723_2241	T/C	0.529	−0.187	−0.71
	126785	MA_9447489_687	A/C	0.083	−0.429	−10.21
LNC	143628	MA_10428744_29330	C/T	0.219	−0.057	−0.52
EP	16868	MA_15729_40331	G/T	0.542	0.149	0.55
	91242	MA_246125_1213	G/A	0.183	−0.129	−1.40
TP	101203	MA_462319_4322	A/C	0.469	−0.199	−0.85
	132014	MA_10251995_2442	A/C	0.339	−0.429	−1.63
LP	162397	MA_10434007_77578	C/T	0.127	−0.071	−1.11
EP/LP	51657	MA_80954_29644	G/A	0.081	0.062	1.49
	60787	MA_98424_947	C/T	0.254	−0.181	−1.43
	123639	MA_8790100_1384	A/C	0.032	−0.078	−4.81
	59480	MA_96191_7122	A/G	0.120	−0.117	−1.95
	117333	MA_1045136_4310	T/C	0.018	0.077	8.56
	166235	MA_10435002_4986	G/A	0.138	0.013	0.19
MI	61096	MA_99004_17108	G/A	0.006	−0.009	−3.16
	67181	MA_109804_10278	G/A	0.007	−0.012	−3.14
	1401	MA_1378_4718	C/A	0.003	0.004	2.67
	138744	MA_10427214_13968	G/T	0.002	0.017	17.00
	162397	MA_10434007_77578	C/T	0.025	−0.010	−0.79
	21924	MA_19222_1789	A/G	0.014	−0.008	−1.14

[a]Calculated as the difference between the phenotype means observed within each homozygous class ($2a = |G_{BB} − G_{bb}|$, where G_{ij} is the trait mean in the ijth genotype class).
[b]Calculated as the difference between the phenotypic mean observed within the heterozygous class and the average phenotypic mean across both homozygous classes [$d = G_{Bb} − 0.5(G_{BB}+G_{bb})$], where G$ij$ is the trait mean in the ijth genotypic class.

Previous work utilizing a functional mapping analysis in forest trees have used a limited number of molecular markers (Li *et al.*, 2014). Li *et al.* (2014) applied this analysis in a bi-parental Scots pine cross using 319 markers. Hence, our work represents an advance in that we have been able to apply this approach at the genome-wide scale (178 101 SNPs) on maternal trees, with a dynamic trait dataset comprising 14 time points/annual growth rings (i.e. cambial age). Latent traits represent significant time points in the trait development allowing us to detect putative genes at these critical junctures in wood formation. Functional mapping has also been applied in ecological studies (Paine *et al.*, 2012) and crops more recently. The fitting of growth models to the data describing growth trajectories of wood

formation phenotypes allowed the identification of marker-trait associations. This enabled us to track phenotype development against the genetic contributions at key time points.

Wood properties have previously been indicated to have a complex genetic architecture, in which association studies that make use of historical recombination represent a method that presents a substantial increase in QTL detection power for such complex traits (Hall *et al.*, 2016). In our study, the number of QTLs detected reflected the complex nature of the traits under study, and our experimental design allowed the detection of the largest/most significant QTLs. A previous functional mapping study involving SNPs in conifers applied two levels of evaluating QTLs (Li *et al.*, 2014), for which they have suggestive and significant QTLs, with our study only reporting the significant QTLs (single level), hence the small number of QTLs in our study. The small number of significant QTLs might also be due to the complex nature of the approximately 20 Gbp spruce genome. The sequence capture method only covered a total of 2331.1 kbp of exonic sequence, 2470.9 kbp of intronic sequence, 40.7 kbp of UTR-like sequence and 9119 exon−intron boundaries (Vidalis *et al.*, 2018). Therefore, a large portion of the genome was not represented and this would be compounded by the rapid LD in spruce, which might affect the number of significant QTLs detected. However, the numbers of QTLs detected in our study are in line with some previous studies in conifers (González-Martínez *et al.*, 2007), and with the drought association study in Norway spruce (29 significant SNP) (Trujillo-Moya *et al.*, 2018).

The QTL detected in our study explain a small proportion of the genetic variation and this could be due to several factors. This is in line with previous studies examining genetic variation in complex traits in coniferous species using forward genetic approaches, such as QTL (Sewell *et al.*, 2000; Novaes *et al.*, 2009) and AM (Wegrzyn *et al.*, 2010; Du *et al.*, 2013, 2018; Porth *et al.*, 2013; McKown *et al.*, 2014; Lamara *et al.*, 2016) The large effective population size in forest tree populations closely resembles humans, therefore making the 'missing heritability' issue found in human AM experiments relevant to forest tree populations. First, one of the hypothesis attributed to this 'missing heritability' is the substantial amount of quantitative variation linked to the cumulative effect of rare alleles that cannot be detected in GWAS using small sample sizes. Therefore in our study increasing the sample size from 517 individuals might allow the inclusion of rare alleles, explaining some of the missing heritability (Hamblin *et al.*, 2011; De La Torre *et al.*, 2019). The detection of true low-frequency alleles associated with complex traits is challenging as it requires large and genetically diverse populations (Hall *et al.*, 2016). Variants with low minor allele frequencies are usually discarded due to the potential of genotyping errors. However, rare alleles play an important

role in both the genetic regulation of traits and explaining the 'missing heritability' in forest species (De La Torre *et al.*, 2019). Therefore, this could have also contributed to the small effect sizes detected in our study as we filtered SNPs with low minor allele frequencies (<0.05 MAF). Second, allelic heterogeneity in which multiple functional alleles exist and are associated with different phenotypes, especially for such complex traits as those linked with wood formation. The presence of allelic heterogeneity will require a large population size that will encompass the allelic variations to account for the missing heritability (Bergelson and Roux, 2010). Third, non-additive effects mainly epistatically derived variation between genes might go undiscovered (Storey *et al.*, 2005). Most GWAS models have been designed to only consider the additive effects of markers. Numerous studies have shown that non-additive effects constitute a large part of the genetic variation of complex traits, these studies considered intra-locus (dominance) and inter-locus (epistatic) effects (Huang *et al.*, 2012; Zhou *et al.*, 2012; Mackay, 2013; Yang *et al.*, 2014; Du *et al.*, 2015). Yang *et al.* (2014) showed in corn an increase in the proportion of heritability explained when a model considering dominance was utilized and therefore allowing a better overview of heterosis. In rice Zhou *et al.* (2012) demonstrated the accumulation of multiple effects, including dominance and overdominance, which might partially explain some of the genetic basis for heterosis. Du *et al.* (2015) identified additive, dominant and epistatic effects explaining nearly two-fold high heritability in *Populus tomentosa* for 10 growth and wood property traits utilizing pathway-based multiple gene associations.

Lastly, epigenetic variation is also likely to be one of the sources of the 'missing heritability'. With the development of advanced sequencing platforms, sophisticated genotyping tools have been developed to unravel epigenetic variation (Johannes *et al.*, 2009). Therefore, the influence of each of these factors on heritability strongly depends on the population sampled and inclusion of sophisticated genotyping tools in the case of epigenetics. The incorporation of a combination of advanced statistical models such as regional heritability mapping (RHM) and the detection of structural variants, insertion/deletions (InDels) and copy number variants in GWAS studies from several tree species has resulted in higher heritabilities being detected (Resende *et al.*, 2017a; Gong *et al.*, 2018)

Trait trajectories and functional mapping

EBVs were plotted as phenotype data versus 14 consecutive cambial ages (Figure 1). All phenotypes under investigation are represented with thin light blue (Black average) curves, to visualize the nature of variation and growth trajectories of the phenotype (Figure 1 and Supporting Information Figure S1). The dissection of dynamic traits in forest trees has been predominantly performed using

(a)

(b)

(c)

(d)
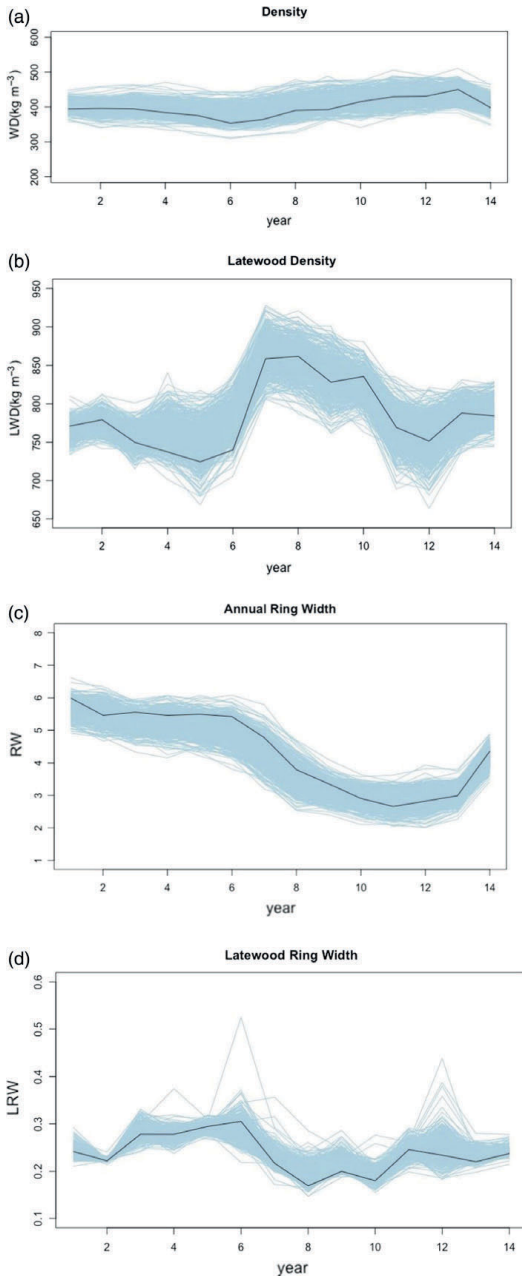
Figure 1. EBV trajectories of four wood quality traits over time: (a) wood density, (b) late wood density, (c) annual ring width and (d) late wood ring width. Individual trajectories for each trait are shown in light blue lines and the black line represents the mean trajectory for the phenotype. These trajectories were used to determine the four latent traits of each tree, using linear splines with two knots.

single data points representing the value of the trait at a given developmental stage. The major disadvantage of such an approach is that it overlooks many of the factors that define the process of formation and development for important traits such as density and ring width. We utilized splines that have the advantage of not making *a priori* assumption about the shape of the curve and allow for the trait growth trends to be unbiased. Splines also allow for the characterization of the dynamic traits in terms of a few parameters derived from the spline models (Al-Tamimi *et al.*, 2016). The fitting of growth trajectories is considered as optimal because it treats phenotypes measured over time as different traits and also takes into account the correlation generated by the ordered time points (Yang *et al.*, 2006). The growth trajectories of the traits over time were calculated from the fitted splines and time intervals were identified and selected based on the characteristic growth trajectory of each trait, resulting in associations *across* and *within* traits being identified (Table 1). Therefore, indicating the control by different sets of genes at different time points for our longitudinal traits (Table 2), just as in some age-specific QTLs found in other conifers and rice (Verhaegen *et al.*, 1997; Emebiri *et al.*, 1998; Wu *et al.*, 1999). This approach has the potential to be applied to genomic prediction and selection studies for predicting individuals that would have the highest impact through the formation and development of a trait of interest. With application of differentially penalized regression (DiPR), pooled significant association markers can be utilized in GS in order to increase prediction accuracies (Bentley *et al.*, 2014).

Linkage disequilibrium

The zygotic LD (squared correlation coefficient r^2) was determined through the pooling of all r^2 values and plotting them against the physical distances between the same SNP pair (Figure 2a). This allowed us to estimate the genome-wide degree of LD in Norway spruce, with the average LD for linked SNPs being inferred from the trendline (curve) of the nonlinear regressions. The fitted curve indicates the LD is low in Norway spruce, rapidly decaying by over 50% (from 0.50 to 0.20) (Figure 2a). The average distance associated with the LD decline for $r^2 = 0.1$ varied from 14 to 1500 bp (Figures 2c,d and S2). Neale and Savolainen (2004) reported an LD decayed to less than 0.20 within roughly 1500 bp based on 19 candidate genes in loblolly pine. As conifers are highly outcrossing a rapid LD decay is expected, however in spruces the LD displays diverse patterns among different genes or the same genes in different species. The LD decline in spruces was also noted to be roughly between a few base pairs and 2000 bp (Namroud *et al.*, 2010). These diverse heterogenous LD patterns were also observed when we analyzed the LD for individual contigs that had significant associations to our traits (Figure 2c,d, Figure S3). The general LD estimate of

all the SNP pairs indicated a fast LD decay (Figure 2a). This rapid decay could be due to the number of contigs analyzed in relation to the large Norway spruce genome, as well as the use of zygotic LD between genotypes. Lu *et al.* (2016), noted that the calculation of gametic LD from phased haplotypes indicated a slower LD decay than when using zygotic LD in loblolly pine. However, they also observed varying rates of LD decay between genes and across different genome regions (Lu *et al.*, 2016). Therefore, the generality of the LD patterns within the Norway spruce genome remains to be further analyzed because only a relatively small and highly specific portions of the genome was studied here.

Population inference

To account for effects derived from population stratification we performed a principal component analysis (PCA). The top two explained a total of 5.3% of the variation. Population structure inference of clusters detected by PCA was performed by ADMIXTURE (Figure S3) and the best K value plotted from the cross-validation error term. Using the best K method, K = 2 better explained the genetic structure of the study population (Figure S3).

Overall summary of genetic associations

Several associations were shared *within* each trait and *across* traits in the analysis. WD, Ring width (RW), Transitional ring widh (TRW) and Latewood number of cells (LNC) had one (MA_33109_11804), two (MA_10434805_21408 and MA_20322_28351), one (MA_817099_1105) and one (MA_10428744_29330) QTLs shared by two or more latent traits, respectively. Common QTLs *within* RW were observed for slope, β_2 and β_3 latent traits, with moderate frequencies ranging from 0.521 to 0.615 and influenced their respective traits to modest degrees (PVE in ranges of 0.18–2.66%).

For QTLs common *across* the different latent traits, SNP MA_10434805_21408 was shared between latewood wood density (LWD), RW and Modulus of elasticity (MOE); this is not surprising because of the close correlation between MOE and wood density. Intron variant MA_10434805_21408 explained between 0.18 and 2.66% of the PVE observed in the respective traits. This SNP associated also had high frequencies of 0.602 and 0.615 in MOE and RW explaining PVE of 1.00 and 2.66%, respectively (Table 2). SNP MA_10435406_13733 was shared between WD, Earlywood wood density (EWD) and Earlywood number of cells (ENC), was associated with the intercept trait for WD and EWD and the slope latent trait in ENC (Table 1), with PVE ranging from 0.01 to 4.64%. The QTL had a high influence on the density related traits as it explained 4.64% (WD) and 3.38% (EWD) both exhibiting a partial dominant inheritance pattern (Table 2).

Numbers of cells (NC), ENC, TNC and LNC traits were associated with a total of three putative genes and three
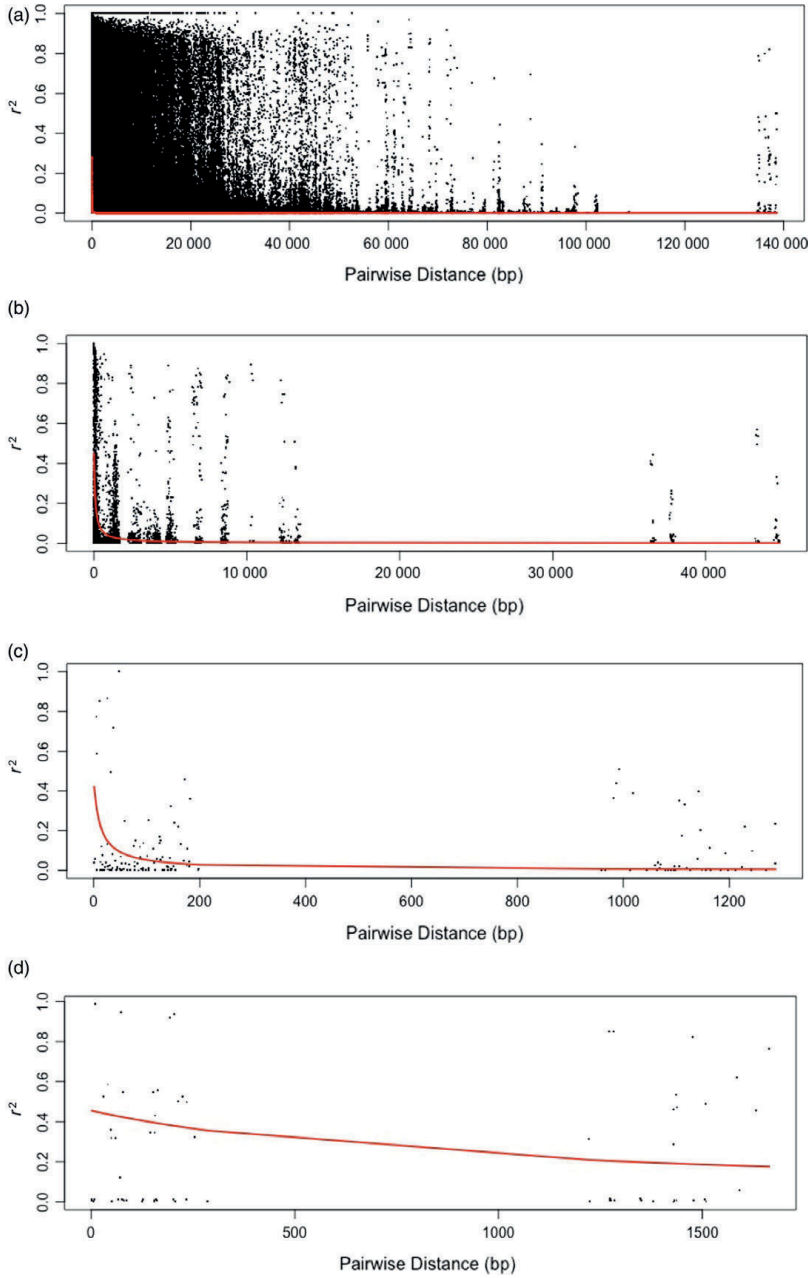
Figure 2. (a) Decay of linkage disequilibrium (LD) across all the tagged genomic sequences, the majority being exonic regions. The squared correlation coefficient between loci (r²) is plotted against distance, in base pairs, separating loci. The fitted curve (red) is representative of the trend of decay from the 178 101 SNPs utilized in the association mapping (AM). (b) Decay of LD with distance in base pairs between sites from across 41 contigs with significant associations. (c) Decay of LD across contig MA_96191 that has a significant association for ratio of percentage earlywood vs latewood on which two probes were captured. (d) Decay of LD on contig MA_80033 indicating the variable LD in the genome.

protein domains. Of the three putative genes, two are associated with serine/kinase activity and one is involved in cysteine and methionine synthesis (Table S1). All the SNPs associated with these traits were either downstream or upstream of coding regions and may therefore act as modifiers of gene expression.

Wood percentage traits, early wood proportion (EP), LP, TP and the ratio of EP/LP had significant associations with 10 SNPs. Four of the six significant SNP variants for EP/LP are modifiers with the other two SNPs, being a synonymous (MA_96191_7122) and missense (MA_1045136_4310) variant. The synonymous SNP MA_96191_7122 was consistent with the codominant mode of inheritance (Table 2). The significant SNP MA_15729_40331, an intron variant, that is associated with EP, is located in the gene MA_15729g0010, homologous to a DNA-3-methyladenine glycosylase II enzyme (Table 1).

WD, EWD, TWD and LWD had a total of 12 significant associations. A missense SNP, MA_33109_11804, was associated with WD and located within the gene homologous to an Arabidopsis senescence-associated gene 24 (Table S1). Of the three significant SNP associations for Transitional wood density (TWD), two, SNP MA_10235390_3386 (stop gained) and SNP MA_212523_89044 (upstream gene variant) were identified within genes. Two of the three significant SNPs identified for LWD were intron variants (MA_10434805_21408 and MA_10436058_4902) with the third being a missense variant (MA_62987_13474).

Trees showing a positive correlation between growth and density had seven QTL specific for this observed phenomenon (MI), explaining a PVE ranging between 0.05 and 1.82% (Table 1). The seven associated SNPs, were two upstream gene variants, two missense variants one intergenic variant, one stop gained variant and one synonymous nucleotide replacement (Table 1). The SNP MA_1378g0010_4718 encodes for a premature stop codon on gene MA_1378g0010. Two SNPs associated with the slope latent trait for MI (MA_1378_4718 and MA_10427214_13968) have an overdominance inheritance pattern with the C and G alleles being dominant, respectively (Table 2; Figure 3).

Genetics associations for genes of known function in wood formation

Intercept associations. Our study identified several interesting genes linked to the significant QTLs from the intercept latent trait, which represents the mean from our spline model. This resulted in 17 significant associations with a

PVE ranging from 0.50 to 4.64% associated with the intercept latent trait. The modes of action determined by the non-additive effects of these significant SNP associations to the intercept latent trait were one for overdominance (|d/a| > 1.25), codominant (|d/a| < 0.50) 12 and four SNPs were partial to fully dominant (0.50 < |d/a| < 1.25).

Ring width phenotypes RW, TRW, and LRW were linked with a total of three gene models associated with the intercept latent trait (Table S1). Of these putative genes associated with RW phenotypes, gene MA_10694g0010 was of particular interest with regards to wood formation. SNP MA_10694_9101 with a partial to fully dominant mode of inheritance (Table 2) was located on the gene MA_10694g0010 that is homologous to an enzyme involved in cell wall biosynthesis, endoglucanase 11-like, and was associated with RW (Table S1) and was expressed in the wood (phloem+cambium+xylem) component of spruce (Figure 4). This enzyme is a vital component of xylogenesis and is involved in the active digestion of the primary cell wall (Goulao *et al.*, 2011). Endoglucanases have been proposed as enzymes involved in controlling cell wall loosening (Cosgrove, 2005). Endoglucanase 11-like is part of the endo-1 family in which the eno-1-4-β-glucanase *Korrigan* gene belongs. Characterization of the *Korrigan* gene in *P. glauca* has identified it to be functionally conserved and essential for cellulose synthesis (Maloney *et al.*, 2012).

Density-related phenotypes (WD, EWD, TWD and LWD) had two significant associations detected for the intercept. Both associations were detected by SNP MA_10435406_13733 and were both partial to fully dominant in their form of inheritance (Table 2). The SNP MA_10435406_13733 downstream on gene MA_10435406g0010 was also significantly associated with the trait ENC slope latent trait. The association of this gene with the WD and EWD intercept implies that it has an impact on the overall development of density throughout the growth period. This result coincides with previous report about the influence of the earlywood component on the properties of the annual ring as a whole in Scots pine (Li *et al.*, 2014). Association of the same gene also with the slope latent trait of ENC corroborates the predictive value of number of cells for wood density. The gene is homologous to phosphoadenosine phosphosulfate reductase (PAPS), which plays a central role in the reduction of sulfur in plants. An analysis of PAPS enzymes in Arabidopsis (Klein and Papenbrock, 2004) and *Populus* (Kopriva *et al.*, 2004) revealed that enzymes involved in sulfate conjugation play an important role in plant growth and development (Klein and Papenbrock, 2004).
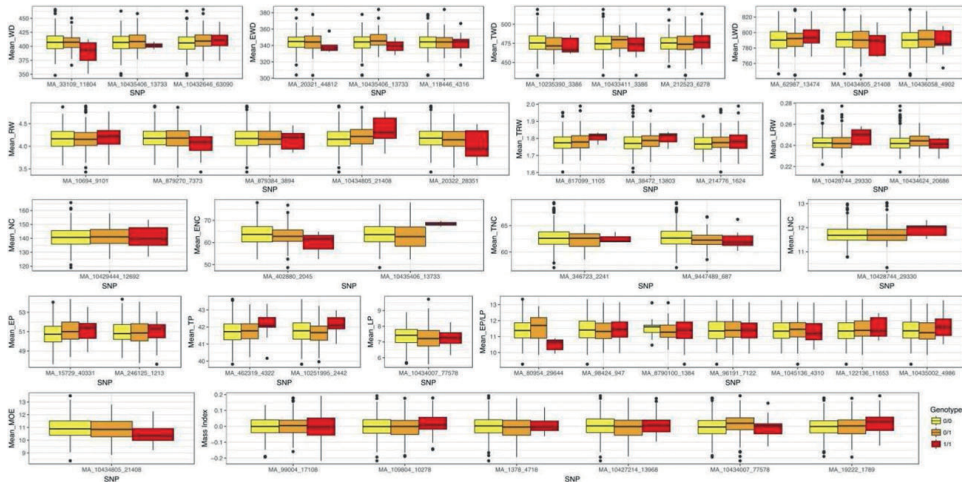
Figure 3. Box plot of the estimated genotypic effect on the phenotypes in the study. The significant SNPs associated and each one of the traits have been correlated to give the impact each genotype has on the average of the overall trait

Slope associations. The analysis of slope (rate of change) of wood formation traits over cambial ages, to our knowledge, has never been dissected in any Norway spruce QTL or AM analyses.

For slope latent traits two significant candidate genes concerning wood formation, PAPS and Proliferating Cell Nuclear Antigen (PCNA), were detected *across* related traits density, growth number of cells and MOE. The codominant SNP MA_10435406_13733 (Table 2) that is a 3′-gene variant for WD was located on a gene that is homologous to a phosphoadenosine phosphosulfate reductase gene cysH_2 and common *across* ENC, WD and EWD.

The SNP MA_10434805_21408 was located on the gene MA_10434805g0010, which is homologous to an Arabidopsis PCNA protein (Table S1) and was ubiquitously expressed with high levels in the wood (phloem+cambium+xylem) component of spruce (Figure 4). This SNP is associated *across* LWD, RW and MOE with partial to fully dominance ($0.50 < |d/a| < 1.25$) for all three associations (Table 2). The presence of these common QTL suggests that these traits might be under the control of the same genes or genetic pathways. Chen *et al.* (2014) reported a significant positive genetic correlation between wood density and MOE, which increased with tree age. However, wood volume growth and density have a negative correlation (Chen *et al.*, 2014), our study was able to detect QTLs for trees exhibiting a positive correlation for this phenomenon (MI). The common QTL observed *across* WD, EWD and ENC indicates that the number of cells during the

juvenile wood development stages has a significant impact on the overall density. The seasonal changes in EWD to LWD have been speculated to be due to a change in auxin levels leading to the initiation of wall-thickening phase, which has a direct impact on the wood quality traits such as MOE. This phase coincides with the cessation of height growth and where available resources are used for cell wall thickening (Sewell *et al.*, 2000), which may explain the common QTL between LWD, RW and MOE, as part of the same feedback loop mechanism.

β_2 and β_3. When analyzing QTLs associated with the two latent traits β_2 and β_3, 16 significant associations were detected, with phenotypic variances ranging from 0.01 to 4.51% (Table 1). Five of the significant markers were consistent with overdominance ($|d/a| > 1.25$), with the 11 markers being dominant ($0.50 < |d/a| < 1.25$) (Table 2, Figure 3).

Wood density phenotypes (WD, EWD, TWD and LWD) had three significant associations with the β_2 and β_3 latent traits (Table 1). The upstream variant MA_33109_11804 associated with both the slope and β_2 latent traits of WD was detected on the gene model MA_33109g0100 that is homologous to the *Arabidopsis* senescence-associated gene 24 (Table S1) and the gene being expressed in shoots and buds of spruce (Figure 4). An association for the latent trait β_2 of TWD with a codominant SNP (MA_212523_6278) (Table 2) was located upstream of gene MA_212523g0010 homologous to Kinesin-related protein 13 (gene-L484_021891) and ubiquitously expressed in shoot, buds and wood component of spruce, indicating its important

Phenotype	Trait	gene_id	confidence	Ath_Nmae/description	Immature male cone	Immature female cone	Needles (2009)	Needles (2008)	Infected needles	Vegetative shoots, May 27, 2010	Vegetative shoots, June 21, 2010	Pineapple galls	Buds early season developing	Needles from vegetative shoots Aug.	Stem from vegetative shoots Aug.	Needles from vegetative shoots Sept.	Buds late season developed	Needles from dried twig	Needles from girdled twig	Stem from girdled twig	Early morning needles 5.30	Mid-day needles 12.00	Late afternoon needles 19.30	Night needles 23.30	Wood (phloem+cambium+xylem early)	Wood (phloem+cambium+xylem late)
Width	RW	MA_879270g0010	High	Alpha-DOX1	−2	1.3	1.1	0	−1	4.6	5.3	4.9	−2.1	1.5	−2	1.5	−1	1	0.2	−2	−1	−0	0.3	−1	−2	−2
		MA_10694g0010	High	GH9B1	3	0	0	0	0	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0	1	0
	RW	MA_10434805g0010	Medium	PCNA1	1.6	−0	−1	−1	−1	−1	−1	0.4	0.9	0.4	0.2	0.7	−0	0	0.9	0	−0	−0	1.2	−1	1	1
		MA_20322g0010	Medium	AP2/B3-like	2.8	−1	−2	0	0.1	−1	0.8	1.7	2.8	0.4	−2	0.5	0.2	−2	−0	−2	−1	−0	−2		3	2
		MA_879384g0010	Medium		0.2	−2	−1	−0	0.3	−2	0.4	1.8	2	0	−2	1.3	−0	−1	−1	−1	−1	−2	1.4	0	0	1
	TRW	MA_214776g0010	Medium	Kinase	−3	1	−0	−0	1.3	0	0.4	−1	−0.8	−2	0.7	−2	2.1	2	−2	0.9	1.1	1.6	−0	1	−4	−5
		MA_38472g0010	High	AHDP	−3	1	0.7	1	0.7	0.9	0.5	−1	−0.4	−1	0.2	−1	−1	−0	−1	0.6	−0	0.2	0.4	−0	−4	−2
		MA_817099g0010	Medium	Senescence protein	0.8	−1	0.1	−1	−1	−0	−0	0.3	0.6	1.8	−0	−1	−1	0	0.7	−1	0.5	0.1	−2	−0	2	1
	LRW	MA_10434624g0010	Medium	PME2	3.9	−1	−1	2	1	1	−3	−3	−2.8	−1	0.2	−3	2.3	1	0.7	−1	−0	0.8	−1	1	4	3
Number of Cells	NC	MA_528074g0010	High	EFR	−2	0.5	0.5	0	0.3	0.7	0	−2	−4	−2	0.8	−3	−2	0	−2	0.2	1.2	0.6	−2	0	−2	−1
	ENC	MA_402880g0010	High	MEE55	1	−1	−0	−1	1.4	−1	2.3	1.4	0.8	2.1	0	2.7	−1	−1	1.6	0	−1	−1	0	−1	−1	0
		MA_10435406g0010	Medium		−0	0.5	−0	0	0.3	−0	−0	−0.2	−1	−1	−1	0.2	−1	1	0.2	−1	0.3	0.7	−0	−1	0	0
	TNC	MA_9447489g0010	Medium	Metallocarboxpeptidase	5.2	3.9	3.5	4	4.1	3.6	4.3	4.1	4.1	3.7	4.9	3.9	3.2	3	4.1	3.2	3.7	3.8	4.1	3	4	4
	LNC	MA_10428744g0010	High	CLK4	−0	0.8	0.1	−0	0.4	−0	0.2	−0	−0.2	−0	−0	−0	0	−0	0.6	0.4	0.6	−1	0	−0	0	
Wood percentage	EP	MA_246125g0010	Medium	TIR-NBS-LRR class	0.1	2.1	0.7	0	0.2	0.4	−0	−2	−1.7	−1	−2	−2	0.5	1	0.4	−0	0.2	0.6	−2	−1	−2	−0
	LP	MA_10434007g0010	High	URM9	0.4	−1	−1	−0	−1	−1	−0	0.1	0.4	−0	−0	0.3	0.5	−0	0.2	−0	−1	−0	0.3	0	1	1
	TP	MA_462319g0010	High	Myosin-binding protein	−0	0.8	0.5	0	0.4	0.6	0.6	−0	−0.2	−0	0.1	−1	0.6	0	−2	−0	0.5	0.1	−0	−0	−0	−0
		MA_80954g0010	Medium	PKS4	−2	−0	−2	−2	3.6	−2	4.3	2.9	−0.4	2.5	1.7	1.3	−2	−0	0.1	−2	−1	0.3	1.4	0	0	−2
		MA_98424g0010	High	STN1	−0	−1	−1	−0	0.2	0.2	0.1	0	0.3	0.8	−0	−0	0.3	0	−1	0.5	0	−0	−0	−0	1	2
	EP/LP	MA_8790100g0010	Medium	AAA-type protein	−0	0.2	−0	0	−0	−0	−0	−1.1	0.3	0.5	−0	0.9	0	0.8	0.5	0	−0	−1	0	0	−0	−0
		MA_96191g0010	Medium	UGT73B2	−2	1.5	0.5	−1	5.3	1.1	4.9	−2	1.3	−2	1.5	−1	0.7	0	−2	−1	−0	0.3	4.6	−1	−2	−2
		MA_460877g0010	Medium	TIR-NBS-LRR class	−0	−2	−0	0.1	−1	0.3	−1	0.7	0.7	−0	−1	0.5	0.5	−0	−0	−1	0	−1	−1	0	0	
Density	Ring	MA_33109g0010	Medium	SAG24	0.5	−1	−1	−0	0.1	−0	−0	−0	0.7	0	0	0.5	0.8	−0	−1	−0	−0.1	−0	0.6	−0	−0	−0
		MA_10432646g0010	High	coil domain	0.2	−0	−0	0	0.2	−0	−0	−0	−0.3	−0	−0.3	−0	0.1	0	0.3	0.4	0	0.1	−0	−0	−0	−0
	EWD	MA_20321g0010	High	BSL2	0	−0	−0	0	−0	0.4	−0	−0	−0.2	−0	0.5	−0	0.5	0	−0	−0	0.6	0.2	−1	1	0	0
		MA_10235390g0010	Medium	MSH5	1.3	0	0	0	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0	0	0	0
	TWD	MA_10433411g0010	Medium	RACK1A	0.1	−1	−0	−1	0	−0	−1	0.3	0.5	−0	−0	0.3	−0	1	0.2	−1	0.2	−0	1	−0	0	0
		MA_21252g0010	High		3.1	−0	−1	−1	−1	−1	0.1	1.7	2	1.4	−1	2.4	−2	−1	1.9	−1	−0	2	−2	3.1	−0	3	2
	LWD	MA_10436058g0010	Medium	Spc97/Spc98 family	0	0	0	0	0	1.2	0	0	1.2	0	0	0	0	0	0	1.8	1	1	0	0	0	0
Mass index	MI	MA_10435002g0010	Medium	OVA2	−1	0.3	0.5	1	−1	−0	−1	−1	−1	−1	0.8	−1	0	1	−1	0.6	1.2	1.4	0.2	1	−2	−2
		MA_99004g0010	Medium	SHR	−1	−0	0.5	0	1	0.1	1.3	−0	1.9	0	−0	0.1	−0	−1	1.2	−1	−0	−1	2.1	−1	−5	−3
		MA_109804g0010	High	YGL	−2	0.4	−0	0	0	−0	0.5	−3	−3.8	−1	1	−4	2.5	0	−0	0.8	0.9	0.5	−2	1	−5	−5
		MA_10427214g0010	Medium		2.9	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4. The heatmap showing the expression levels (VST values) of spruce candidate genes in different organs and tissues based on data of Nystedt *et al.* (2013) available at http://congenie.org.

function in this species (Figure 4). This association is of interest because Kinesin-related proteins are known to be involved in secondary wall deposition, which can impact wood density (Zhong *et al.*, 2002), cell wall strength and oriented deposition of cellulose microfibrils. Therefore, these proteins would have a direct correlation with the increase in density at the latent trait β_2 at age 6 years (Figure S1).

Ring width phenotypes (RW, TRW and LRW) had eight significant associations identified for the latent traits β_2 and β_3, explaining PVE ranging from 0.01 to 4.51% (Table 1). The synonymous SNP MA_2032_28351 associated with the β_2 latent trait for RW is located on a gene homologous with a plant-specific B3-DNA binding protein domain explaining 1.78% variation and is shared among various plant-specific transcription factors. This includes transcription factors involved in auxin and abscisic acid responsive transcription (Yamasaki *et al.*, 2004). Auxin is one of the central phytohormones in the control of plant growth and development (Abel and Theologis, 1996), and also known to be involved in cell wall loosening and

elongation (Cosgrove, 2016). This association was detected *within* the RW phenotype and detected for both β_2 and β_3 latent traits (Table 1). Therefore, this domain could be involved with transcription factors involved in both the decrease and increase of RW (Figure S1). Three putative genes were associated with the β_2 and β_3 latent traits for the TRW phenotype. Of interest a senescence-associated protein associated on the TRW β_2 latent trait with the missense variant MA_817099_1105. This might be linked to the decrease in TRW at year 6 (Figure S1) due to the decline of photosynthetic rate known to be induced by the activity of senescence related proteins (Sillanpää *et al.*, 2005). The gene was highly expressed in both the early and late wood components of spruce supporting the row of these senescence genes in controlling tree growth (Figure 4). This association was also identified for the slope latent trait indicating a potential impact on the rate of change of transitional wood. The detection of senescence related genes for wood density related phenotypes for both the slope and β_2 latent trait (MA_33109_11804) could indicate a possible relationship between the genes

influencing RW and wood density. Contig MA_10434624 is homologous to a pectin esterase and was associated with the downstream variant MA_10434624_20686 with an over-dominant mode of inheritance for LRW (β_3 latent trait) at year 10 (Table 2, Figure S1). This significant downstream SNP (MA_10434624_20686) associated with LRW on gene MA_10434624g0020 and homologous to pectin methylesterases (PMEs), which are cell wall associated enzymes responsible for demethylation of polygalacturonans (Phan *et al.*, 2007). This gene was observed highly expressed in developing wood (Figure 4), indicating its importance for growth in spruce. This enzyme has been shown to be linked with many developmental processes in plants, such as, cellular adhesion and stem elongation (Micheli, 2001). An association study in *Picea glauca* (Moench) Voss identified a significant nonsynonymous SNP coding for cysteine associated with earlywood and total wood cell wall thickness associated with pectin methylesterase (Beaulieu *et al.*, 2011). Our study identified a PME SNP association in the late wood stage, supporting the importance of PMEs in wood cell development.

When analyzing QTLs detected for traits linked to the percentage of cells (EP, LP and EP/LP) we identified three putative candidate genes, DNA-3-methyladenine glycosylase II enzyme, phytochrome kinase substrate 1 and glycosyltransferase. Synonymous SNP (MA_96191_59480) within the gene MA_96191g0010, which is homologous to Glucosyltransferase in *Picea sitchensis* was associated with the EP/LP, β_2 latent trait. The gene is highly expressed in vegetative shoots (June) and during the late afternoon in needles (Figure 4). Glycosyl transferases operate by facilitating the catalytic sequential transfer of sugars from activated donors to acceptor molecules that form region and stereospecific glycosidic linkages (Lairson *et al.*, 2008). The Arabidopsis ortholog (UDP-glucosyltransferase 73B2) encodes for a putative flavonol 7-*O*-glucosyltransferase involved in stress responses. In our study, this significant association was associated with EP/LP, however a nonsynonymous variant in a gene coding for a glycosyl transferase in *Populus* was associated with fibre development and elongation (Porth *et al.*, 2013).

Several receptor-like kinases (TIR/NBS/LRR and serine/threonine-protein phosphatase) homologues were identified *across* traits (TRW, NC, EP, EP/LP and EWD) (Table S1). Approximately 2.5% of the annotated genes in Arabidopsis genome are RLK homologues (Shiu and Bleecker, 2001), in which they, among other functions, play an important role in the differentiation and separation of xylem and phloem cells (Fisher and Turner, 2007). Similar to our study a synonymous SNP in an RLK gene was associated with EP in white spruce (Beaulieu *et al.*, 2011), hence RLKs seem to be involved in modifying a number of different wood properties from density to cell identity and number.

Norway spruce trees that possess the ability of fast growth and high wood density are very rare, but such trees and associated SNPs were discovered in our study. Trees combining these traits are of high interest to the forestry industry. Of the seven genes significantly linked to this phenomenon of particular interest was a synonymous SNP on MA_99004g0100 gene homologous to a transcription factor from the GRAS family (Table S1). GRAS is an important class of plant-specific proteins derived from three members: GIBBERELLIC ACID INSENSITIVE (GAI), REPRESSOR of GAI (RGA) and SCARECROW (SCR) (GRAS) (Hirsch and Oldroyd, 2009). GRAS genes are known to be involved in the regulation of plant development through the regulation of gibberellic acid (GA) and light signalling (Hirsch and Oldroyd, 2009; Cenci and Rouard, 2017). Furthermore GA signalling has also been shown to stimulate wood formation in *Populus* (Mauriat and Moritz, 2009). Therefore, the GRAS transcription factor identified here and the other six genes positively associated with MI provide interesting genetic markers and tools to understand this phenomenon.

Wood density traits were associated with a total of 12 genes, the largest number of genes identified from the contigs. The percentage of wood was linked to 10 putative genes, cell width had nine putative genes and number of cells was associated with six genes. Two genes were shared *across* multiple traits, PCNA was common *across* RW and LWD, and phosphoadenosine phosphosulfate reductase was shared *across* WD, EWD and ENC.

EXPERIMENTAL PROCEDURES

Plant material and phenotype data

Plant material and phenotype data used in this study have previously been described in Chen *et al.* (2014). In brief, two progeny trials were established in 1990 in southern Sweden (S21F9021146 aka F1146 (trial1) and S21F9021147 aka F1147 (trial2)). We selected 517 families originating from 112 sampling stands to use in the investigation of wood properties. At each site, increment wood cores of 12 mm were collected at breast height from six trees of the selected families (1.3 m) (6 progeny × 2 sites = 12 progenies in total). In total, 5618 trees, 2973 and 2645 trees from the F1146 and F1147 trials respectively, were analysed. The pith to bark profiling of growth and wood physical attributes was performed using the SilviScan instrument (Evans and Ilic, 2001) at Innventia, now part of RISE, Stockholm, Sweden, where also the initial data evaluations were performed (Methods S2). These included the identification and dating of all annual rings and their compartments of early wood (EW), transition wood (TW) and late wood (LW). For this, a density-based '20-80' definition was used, described and discussed in (Lundqvist *et al.*, 2018). Traits of interest to breeders were derived from the SilviScan data, such as the radial NC and Mass Index (MI) introduced to express the relative amount of biomass at breast height.

The investigation was trigged by the observation that some trees broke the unfavourable negative correlation of the trait MI which is between density and growth. They produced, more biomass than expected, and it was therefore important. In order to

identify putative genes behind high values for this trait. MI was defined as:

$$\text{Mass index} = (\text{Individual cross} - \text{sectional average density}$$
$$/ \text{population cross} - \text{sectional average density})$$
$$* (\text{individual cross} - \text{sectional area}$$
$$/ \text{population average cross} - \text{sectional area})$$

The traits investigated in this study are listed in Table 3.

Statistical analysis

EBVs were calculated for growth and wood quality traits for 14 consecutive annual growth rings. The variance and covariance components were estimated using ASREML 4.0 (Gilmour *et al.*, 2014) as described in Chen *et al.* (2014). In brief, the EBVs at each cambial age were estimated using univariate, bivariate or multivariate mixed linear models. The following univariate linear mixed model for joint-site analysis was fitted to calculate EBV:

$$Y_{ijkl} = \mu + S_i + B_{j(i)} + F_k + SF_{ik} + e_{ijkl} \qquad (1)$$

where Y_{ijkl} is the observation on the lth tree from the kth family in jth block within the ith site, μ is the general mean, S_i and $B_{j(i)}$ are the fixed effects of the ith site and the jth block within the ith site, respectively, F_k and SF_{ik} are the random effects of the kth family and the random interactive effect of the ith site and kth family, respectively, e_{ijkl} is the random residual effect. The random family and site by family interaction effects are assumed to follow $N(0, \sigma_f^2)$ $N(0, \sigma_{sf}^2)$ and, respectively, where σ_f^2 and σ_{sf}^2 are the estimated family genetic variance and site by family interaction variance, respectively. Residual variation e was assumed to be $N\left(0, \begin{bmatrix} I_{n1}\sigma_{e1}^2 & 0 \\ 0 & I_{n2}\sigma_{e2}^2 \end{bmatrix}\right)$, where $\sigma_{e_1}^2$ and $\sigma_{e_2}^2$ are the residual variances for site 1 and site 2, I_{n1} and I_{n2} are identity matrices, $n1$ and $n2$ are the number of individuals in each site. The fit of different models was evaluated using the Akaike Information Criteria (AIC) and the optimal model was selected based on a compromise of model fit and complexity.

Table 3 List of the phenotypes, their abbreviations and measurement unit

Phenotype	Abbreviation	Unit
Ring wood density	WD	kg m^{-3}
Early wood density	EWD	kg m^{-3}
Transition wood density	TWD	kg m^{-3}
Late wood density	LWD	kg m^{-3}
Ring width	RW	μm
Early wood ring width	ERW	μm
Transition wood ring width	TRW	μm
Late wood ring width	LRW	μm
Ring number of cells	NC	
Early wood number of cells	ENC	
Transition wood number of cells	TNC	
Late wood number of cells	LNC	
Early wood percentage	EP	%
Transition wood percentage	TP	%
Late wood percentage	LP	%
Early/late wood percentage	EP/LP	%
Modulus of elasticity	MOE	GPa
Mass index (density × growth)	MI	

Latent traits

The EBVs were plotted against cambial age (annual ring number) to produce time trajectories for each trait (Figures 1 and S1). Spline model was fitted to the trajectories and their curve parameters describing the character of their development over time were used as latent traits in order to describe the dynamics of the EBVs across age.

The general definition of a linear spline with multiple knots is as follows

$$y(t) = \beta_0 + \beta_1 t + \beta_2(t - K_1)_+ + \beta_3(t - K_2)_+ + \dots$$
$$+ \beta_{1+m}(t - K_m)_+, \qquad (2)$$

which is continuous and where K_i ($i = 1, \dots, m$; $K_1 < K_2 \dots < K_m$) are defined as knots, and $(t - K_i)_+ = (t - K_i)$ if $t > K_i$ ($K_i > 0$; $i = 1, \dots, m$), and otherwise is equal to zero. The number of knots has to be properly defined to provide an accurate description of the data under investigation, while avoiding overadaptation to data (Li and Sillanpää, 2015; Camargo *et al.*, 2018). In our case, we found two knots most suitable to the time intervals investigated. Hence, the linear spline model to describe the growth trajectory of individual i applied in this study was defined as:

$$y(t) = \beta_0 + \beta_1 t + \beta_2(t - K_1)_+ + \beta_3(t - K_2)_+ + \varepsilon_i(t),$$
$$\varepsilon_i(t) \overset{\text{iid}}{\sim} N(0, \sigma^2). \qquad (3)$$

In equation (2), the intercept β_0, slope parameters β_1, β_2 (at Knot 1 (K_1)) and β_3 (at Knot 2 (K_2)) are estimated by standard least squares (Ruppert *et al.*, 2003). The four estimates were used as the latent trait in the subsequent QTL analysis conducted in RSTUDIO (Team, 2015), and then analysed using the LASSO model to identify SNPs showing significant associations to the traits.

The intercept and slopes were used to evaluate the mean and rate of change for the trait across the annual rings, respectively. β_2 and β_3 represent inflection points in the cambial age trajectories where the development of the EBVs enters new phases. These two points (β_2 and β_3) are therefore supposed to have biological significance, warranting a closer analysis of the genes imparting these shifts in the EBVs dynamics. The four latent traits show lower correlations compared with the direct measurements on the original scales and they also have constant variances, thereby reducing the need to account for residual dependencies in the model (Wu *et al.*, 2004; Yang and Xu, 2007; Li *et al.*, 2014).

Sequence capture, genotyping and SNP annotation

Total genomic DNA was extracted from 517 maternal trees, using the Qiagen Plant DNA extraction (Qiagen, Hilden, Germany) protocol with DNA quantification performed using the Qubit® ds DNA Broad Range (BR) Assay Kit (Oregon, USA). Extracted DNA was submitted to RAPiD Genomics (USA) where DNA library preparation and capture sequencing were performed. Sequence capture was performed using the 40 018 diploid probes designed and evaluated for *P. abies* (Vidalis *et al.*, 2018). The Illumina sequencing compatible libraries were amplified with 14 cycles of polymerase chain reaction (PCR) and the probes were then hybridized to a pool comprising 500 ng of 8 equimolar combined libraries following Agilent's SureSelect Target Enrichment System (Agilent Technologies, https://www.agilent.com/). These enriched libraries were then sequenced using an Illumina HiSeq 2500 instrument (San Diego, USA) on the 2 × 100 bp sequencing mode.

Raw reads were mapped against the *P. abies* reference genome v.1.0 using BWA-MEM (Li, 2013). SAMTOOLS v.1.2 (Li *et al.*, 2009) and Picard (http://broadinstitute.github.io/picard) were used for sorting and marking of PCR duplicates. Variant calling was performed using GATK HAPLOTYPECALLER v.3.6 (Van der Auwera *et al.*, 2013) in gVCF output format. Samples were then merged into batches of approximately 200 before all 517 samples were jointly called.

Variant Quality Score Recalibration (VQSR) method was performed to avoid the use of hard filtering for exome/sequence capture data. For the VQSR analysis two datasets were created, a training subset and an input file. The training dataset was derived from the Norway spruce genetic mapping population showing expected segregation patterns (Bernhardsson *et al.*, 2019) and assigned a prior value of 15.0. The input file was derived from the raw sequence data using GATK with the following parameters: extended probe coordinates by +100 excluding INDELS, excluding LowQual sites, and keeping only bi-allelic sites. The following annotation parameters QualByDepth (QD), MappingQuality (MQ) and BaseQRankSum, with tranches 100, 99.9, 99.0 and 90.0 were then applied for the determination of the good versus bad variant annotation profiles. After obtaining the variant annotation profiles, the recalibration was then applied to filter the raw variants. Using VCFTOOLS v.0.1.13 (Danecek *et al.*, 2011), SNP trimming and cleaning involved the removal of any SNP with a MAF and 'missingness' of <0.05 and >20%, respectively. The resultant SNPs were annotated using default parameters for SNPEFF 4 (Cingolani *et al.*, 2012). Ensembl general feature format (GFF; gene sets) information was utilized to build the *P. abies* SNPEFF database.

Genetic structure and mode of inheritance

Linkage disequilibrium was calculated as the squared correlation coefficient between genotypes (r^2), globally with special attention given to all the contigs with significant associations in VCFTOOLS v.0.1.13 software using the 'geno-r2' routine (Danecek *et al.*, 2011). The trendline of LD decay with physical distance was fitted using nonlinear regression (Hill and Weir, 1988) and the regression line was displayed using RSTUDIO (Team, 2015). Non-additive effects of the significant markers was determined using the ratio of dominance (*d*) to additive (*a*). The ranges were: partial or complete dominance ($-0.50 < |d/a| < 1.25$) and additive ($-0.50 \leq |d/a| \leq 0.50$), with $|d/a| > 1.25$ being equal to over- or underdominance (Eckert *et al.*, 2009).

FactoMiner (Multivariate Exploratory Data Analysis and Data Mining) (Husson *et al.*, 2017) implemented in RSTUDIO software was used to perform PCA. The covariate matrix derived from the PCA was then displayed by plotting principal component 1 scores against principal component 2 scores. The components of the PCA covariate matrix were then applied to the AM to account for population structure and correcting for any stratification within the study. Significance of each genetic principal component (PC) was determined using the Tracy-Widom (TWi) distribution and a significance threshold of $P = 0.01$. For population clustering, ADMIXTURE v.1.3.0 (Alexander *et al.*, 2009) was used with five-fold cross-validation and 200 bootstrap replicates. The bestK method was implemented in RSTUDIO to determine the best K with the use of an elbow plot on the cross-validation error.

Trait association mapping

It is natural to use LASSO method for simultaneous estimation of SNP effects and selecting a sparse subset of trait-associated SNPs to the multilocus association model. This is because LASSO has nice properties like being able to handle high-dimensional cases with p>>n (*i.e.*, a number of SNPs much larger than number of individuals) and selecting only a single representative SNP from the group of highly dependent SNPs. The LASSO model as described by Li *et al.* (2014), was applied to all latent traits for the detection of QTLs.

The LASSO model:

$$\min_{(\alpha_0, \alpha_j)} \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \alpha_0 - \sum_{j=1}^{p} x_{ij}\alpha_j \right)^2 + \lambda \sum_{j=1}^{p} |\alpha_j|, \quad (4)$$

where y_i is the phenotypic value of an individual i ($i = 1,\ldots,n$; n is the total number of individuals) for the latent trait β_0, β_1, β_2 or β_3, α_0 is the population mean parameter, x_{ij} is the genotypic value of individual i and marker j coded as 0, 1 and 2 for three marker genotypes AA, AB and BB, respectively, α_j is the effect of marker j ($i = 1,\ldots,n$; n is the total number of markers), and λ (>0) is a shrinkage tuning parameter. The penalty term is able to shrink the additive effects of some of the markers exactly to zero, and select a subset of the most important markers into the model. The tuning parameter λ determines the degree of shrinkage, and the number of markers having non-zero effects. Cross-validation is used to decide an optimal value for λ.

In *stability selection* (Meinshausen and Bühlmann, 2010; Alexander and Lange, 2011): (i) several bootstrap samples are first created from the original data; and (ii) frequency over-bootstrap samples on how many times each SNP is being selected to the LASSO model is monitored and used as a stable measure of variable selection. Stability selection probability (SSP) of each SNP being selected to the model was applied as a way to control the false discovery rate and determine significant SNPs (Gao *et al.*, 2014; Li and Sillanpää, 2015). Briefly a subsample of half the number of individuals was randomly picked up and the LASSO was performed on it to select a set of markers. This procedure was repeated 1000 times. Then the selection frequency of each marker being selected was calculated, and was used to judge the support of QTL. A decision rule suggested by Meinshausen and Bühlmann (2010) was applied to control the expected number of false positives:

$$\frac{1}{2} + \frac{q^2}{2E[V]p}, \quad (5)$$

where q is the number of selected markers, $E[V]$ is the expected number of false positives, and p is the total number of markers. For a marker to be declared as a significant QTL, a SSP inclusion frequency of at least 0.52 (i.e. derived based on formula 5) was used for all traits. This frequency was inferred conditional on the expected number of false selected markers being less than one (Bühlmann *et al.*, 2014).

Population structure was accounted for in all analyses by including the first five PCs based on the genotype data as covariates into the model. An adaptive LASSO approach (Zou, 2006) was used to determine the percentage of phenotypic variance (PVE) (H^2_{QT}) of all the QTLs (Methods S1). These analysis were all performed in RSTUDIO (Team, 2015).

Candidate gene mining

To assess putative functionality of SNPs with significant associations, a gene enrichment analysis of putative genes and their associated orthologs was performed against the NORWOOD v1.0 database (http://norwood.congenie.org) hosted by CONGENIE (http://congenie.org/). The complete *P. abies* contigs that harboured the QTLs that were not annotated in the CONGENIE were

used to perform a nucleotide BLAST (BLASTN) search, using the option for only highly similar sequences (MEGABLAST) in the National Center for Biotechnology Information (NCBI) nucleotide collection database (https://blast.ncbi.nlm.nih.gov/Blast.cgi?).

CONCLUSION

This work has dissected the genetic basis of wood properties in Norway spruce with use of functional AM. In total, we identified 52 significant QTLs for wood properties and mining of candidate genes located in the vicinity of significant QTLs identified genes that could be directly or indirectly responsible for variations in the observed traits. Functional mapping analyses allowed us to utilize all the longitudinal data for a trait simultaneously and may better account for the temporal trends and correlation structures across years for the complex traits associated with wood formation. It can therefore be applied to the detection of QTLs stable over time (i.e. the QTLs associated with intercept traits) with greater statistical evidence. The slope latent trait over cambial ages or the rate of juvenile-to-mature wood transition has allowed for the dissection the dynamics of the transition process itself and can be applied to other important plant breeding traits. The significance of our results is provided by the identification of QTLs associated to both high wood density and fast growth, therefore larger biomass. These QTLs can now be a basis for future functional genomics in Norway spruce.

However, the direct use of QTLs for marker-assisted breeding has not been successful, mainly due to the difficulty in transferring the associations across populations and species of forest trees. With the small percentage variances detected and no direct information about the developmental change of QTL expression, breeders will be unable to make use of these QTLs in direct early selection. Non-additive interactions especially epistasis, play an important role in accounting for the total genetic variance of a trait. Therefore this study will be a good basis for initiating the detection and estimation of possible epistatic influence on these complex traits. Future work should focus on replicated sampling from a larger number of representative genotypes across different environments, which take into consideration genotype × environment interactions. Additional support for marker-assisted tree breeding may also be provided by the functional genetics studies, systems mapping and consideration of biological mechanisms (Liu and Yan, 2019) of the identified candidate genes in model trees like *Populus* sp.

DATA AVAILABILITY

All the latent traits, genotypic data, SNP position files the AM scripts used for the analysis are publicly available at zenodo.org at https://doi.org/10.5281/zenodo.1480536. Raw sequence data for all the samples utilized in the study are found through the European Nucleotide Archive under accession number PRJEB29652. The Norway spruce genome assemblies and resources are available from http://congenie.org/pabiesgenome.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors declare no competing interests.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Phenotype trajectories representing the main traits.
Figure S2. Significant contigs LD heatmap.
Figure S3. ADMIXTURE plot of the entire population.
Figure S4. Data are structured into three categories.
Methods S1. PVE evaluation of a QTL.
Methods S2. Trait data set used for GWAS identification of novel candidate loci affecting number of tracheids formed, radial growth, density, stiffness and mass at breast height of young Norway spruce.
Table S1. ConGenIE BLAST search of contigs with significant QTLs.
Table S2. Ring-related data (B): list of variables and examples of data.
Table S3. Curve shape data (A): list of variables for each property and example of data.

REFERENCES

Abel, S. and Theologis, A. (1996) Early Genes and Auxin Action. *Plant Physiol.* **111**(1), 9.

Alexander, D.H. and Lange, K. (2011) Stability selection for genome-wide association. *Genet. Epidemiol.* **35**(7), 722–728. https://doi.org/10.1002/gepi.20623.

Alexander, D., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**(9), 1655–1664.

Al-Tamimi, N., Brien, C., Oakey, H., Berger, B., Saade, S., Ho, Y.S., Schmöckel, S.M., Tester, M. and Negrão, S. (2016) Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping. *Nat. Commun.* **7**, 13342. https://doi.org/10.1038/ncomms13342. https://www.nature.com/articles/ncomms13342#supplementary-information.

Anderegg, W.R.L. (2015) Spatial and temporal variation in plant hydraulic traits and their relevance for climate change impacts on vegetation. *New Phytol.* **205**(3), 1008–1014. https://doi.org/doi:10.1111/nph.12907.

Beaulieu, J., Doerksen, T., Boyle, B. *et al.* (2011) Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. *Genetics*, **188**(1), 197–214. https://doi.org/10.1534/genetics.110.125781.

Beavis, W.D. (1998) QTL analyses: power, precision, and accuracy. *Mol Dissect Complex Traits*, **1998**, 145–162.

Bentley, A.R., Scutari, M., Gosman, N. *et al.* (2014) Applying association mapping and genomic selection to the dissection of key traits in elite European wheat. *Theor. Appl. Genet.* **127**(12), 2619–2633. https://doi.org/10.1007/s00122-014-2403-y.

Bergelson, J. and Roux, F. (2010) Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nat. Rev. Genet.* **11**, 867. https://doi.org/10.1038/nrg2896.

Bernhardsson, C., Vidalis, A., Wang, X., Scofield, D.G., Schiffthaler, B., Baison, J., Street, N.R., García-Gil, M.R. and Ingvarsson, P.K. (2019) An ultra-dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway spruce (*Picea abies*). *G3: Genes - Genomes - Genetics*, **9**(5), 1623–1632. g3.200840.202018. https://doi.org/10.1534/g3.118.200840.

Bertaud, F. and Holmbom, B. (2004) Chemical composition of earlywood and latewood in Norway spruce heartwood, sapwood and transition zone wood. *Wood Sci. Technol.* **38**(4), 245–256. https://doi.org/10.1007/s00226-004-0241-9.

Bühlmann, P., Kalisch, M. and Meier, L. (2014) High-dimensional statistics with a view toward applications in biology. *Annu. Rev. Stat. Appl.* **1**(1), 255–278. https://doi.org/10.1146/annurev-statistics-022513-115545.

Camargo, A.V., Mackay, I., Mott, R., Han, J., Doonan, J.H., Askew, K., Corke, F., Williams, K. and Bentley, A.R. (2018) Functional mapping of quantitative trait loci (QTLs) associated with plant performance in a wheat MAGIC mapping population. *Front. Plant Sci.* **9**, 887. https://doi.org/10.3389/fpls.2018.00887.

Cenci, A. and Rouard, M. (2017) Evolutionary analyses of GRAS transcription factors in angiosperms. *Front. Plant Sci.* 8, 273. https://doi.org/10.3389/fpls.2017.00273.

Chen, Z.-Q., Gil, M.R.G., Karlsson, B., Lundqvist, S.-O., Olsson, L. and Wu, H.X. (2014) Inheritance of growth and solid wood quality traits in a large Norway spruce population tested at two locations in southern Sweden. *Tree Genet. Genomes*, 10(5), 1291–1303.

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

Cosgrove, D.J. (2005) Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* 6(11), 850–861. https://doi.org/http://www.nature.com/nrm/journal/v6/n11/suppinfo/nrm1746_S1.html.

Cosgrove, D.J. (2016). Catalysts of plant cell wall loosening. *F1000Res.* 5, F1000. Faculty Rev-1119. https://doi.org/10.12688/f1000research.7180.1

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

De La Torre, A.R., Puiu, D., Crepeau, M.W., Stevens, K., Salzberg, S.L., Langley, C.H. and Neale, D.B. (2019) Genomic architecture of complex traits in loblolly pine. *New Phytol.* 221(4), 1789–1801. https://doi.org/doi:10.1111/nph.15535.

Du, Q., Pan, W., Xu, B., Li, B. and Zhang, D. (2013) Polymorphic simple sequence repeat (SSR) loci within cellulose synthase (PtoCesA) genes are associated with growth and wood properties in Populus tomentosa. *New Phytol.* 197(3), 763–776. https://doi.org/10.1111/nph.12072.

Du, Q., Tian, J., Yang, X., Pan, W., Xu, B., Li, B., Ingvarsson, P.K. and Zhang, D. (2015) Identification of additive, dominant, and epistatic variation conferred by key genes in cellulose biosynthesis pathway in Populus tomentosa†. *DNA Res.* 22(1), 53–67. https://doi.org/10.1093/dnares/dsu040.

Du, Q., Lu, W., Quan, M., Xiao, L., Song, F., Li, P., Zhou, D., Xie, J., Wang, L. and Zhang, D. (2018) Genome-wide association studies to improve wood properties: challenges and prospects. *Front. Plant Sci.* 9, 1912. https://doi.org/10.3389/fpls.2018.01912

Dutilleul, P., Herman, M. and Avella-Shaw, T. (1998) Growth rate effects on correlations among ring width, wood density, and mean tracheid length in Norway spruce (Picea abies). *Can. J. For. Res.* 28(1), 56–68.

Eckert, A.J., Bower, A.D., Wegrzyn, J.L., Pande, B., Jermstad, K.D., Krutovsky, K.V., Clair, J.B.S. and Neale, D.B. (2009) Association genetics of coastal Douglas fir (Pseudotsuga menziesii var. menziesii, Pinaceae). I. Cold-hardiness related traits. *Genetics*, 182(4), 1289–1302.

Emebiri, L.C., Devey, M.E., Matheson, A.C. and Slee, M.U. (1998) Age-related changes in the expression of QTLs for growth in radiata pine seedlings. *Theor. Appl. Genet.* 97(7), 1053–1061. https://doi.org/10.1007/s001220050991.

Evans, R. and Ilic, J. (2001) Rapid prediction of wood stiffness from microfibril angle and density. *Forest Products J.* 51(3), 53.

Fisher, K. and Turner, S. (2007) PXY, a receptor-like kinase essential for maintaining polarity during plant vascular-tissue development. *Curr. Biol.* 17(12), 1061–1066.

Gao, H., Wu, Y., Li, J., Li, H., Li, J. and Yang, R. (2014) Forward LASSO analysis for high-order interactions in genome-wide association study. *Brief. Bioinform.* 15(4), 552–561.

Gilmour, A., Gogel, B., Cullis, B., Welham, S., Thompson, R., Butler, D., Cherry, M., Collins, D., Dutkowski, G. and Harding, S. (2014). ASReml user guide. Release 4.1 structural specification. *VSN International Ltd, Hemel Hempstead, HP1 1ES, UK* www.vsni.co.uk.

Gong, C., Du, Q., Xie, J., Quan, M., Chen, B. and Zhang, D. (2018) Dissection of insertion–deletion variants within differentially expressed genes involved in wood formation in Populus. *Front. Plant Sci.* 8, 2199. https://doi.org/10.3389/fpls.2017.02199.

González-Martínez, S.C., Wheeler, N.C., Ersoz, E., Nelson, C.D. and Neale, D.B. (2007) Association Genetics in Pinus taeda L. I. Wood Property Traits. *Genetics*, 175(1), 399–409. https://doi.org/10.1534/genetics.106.061127.

Goulao, L.F., Vieira-Silva, S. and Jackson, P.A. (2011) Association of hemicellulose- and pectin-modifying gene expression with Eucalyptus globulus secondary growth. *Plant Physiol. Biochem.* 49(8), 873–881. https://doi.org/10.1016/j.plaphy.2011.02.020.

Groover, A., Devey, M., Fiddler, T., Lee, J., Megraw, R., Mitchel-Olds, T., Sherman, B., Vujcic, S., Williams, C. and Neale, D. (1994) Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics*, 138(4), 1293–1300.

Hall, D., Hallingbäck, H.R. and Wu, H.X. (2016) Estimation of number and size of QTL effects in forest tree traits. *Tree Genet. Genomes*, 12(6), 110. https://doi.org/10.1007/s11295-016-1073-0.

Hallingbäck, H.R., Sánchez, L. and Wu, H.X. (2014) Single versus subdivided population strategies in breeding against an adverse genetic correlation. *Tree Genet. Genomes*, 10(3), 605–617.

Hamblin, M.T., Buckler, E.S. and Jannink, J.-L. (2011) Population genetics of genomics-based crop improvement methods. *Trends Genet.* 27(3), 98–106. https://doi.org/10.1016/j.tig.2010.12.003.

Hannrup, B., Cahalan, C., Chantre, G., Grabner, M., Karlsson, B., Bayon, I.L., Jones, G.L., Müller, U., Pereira, H. and Rodrigues, J.C. (2004) Genetic parameters of growth and wood quality traits in Picea abies. *Scand. J. For. Res.* 19(1), 14–29.

Hauksson, J.B., Bergqvist, G., Bergsten, U., Sjöström, M. and Edlund, U. (2001) Prediction of basic wood properties for Norway spruce. Interpretation of Near Infrared Spectroscopy data using partial least squares regression. *Wood Sci. Technol.* 35(6), 475–485. https://doi.org/10.1007/s00226-001-0123-3.

Heuven, H.C. and Janss, L.L. (2010). Bayesian multi-QTL mapping for growth curve parameters. Paper presented at the BMC proceedings, 4(1), S12.

Hill, W. and Weir, B. (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* 33(1), 54–78.

Hirsch, S. and Oldroyd, G.E.D. (2009) GRAS-domain transcription factors that regulate plant development. *Plant Signal. Behav.* 4(8), 698–700.

Hirsch, C.D., Evans, J., Buell, C.R. and Hirsch, C.N. (2014) Reduced representation approaches to interrogate genome diversity in large repetitive plant genomes. *Brief. Funct. Genomics*, 13(4), 257–267.

Huang, X. and Han, B. (2014) Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* 65, 531–551.

Huang, W., Richards, S., Carbone, M.A. et al. (2012) Epistasis dominates the genetic architecture of Drosophila quantitative traits. *Proc. Natl Acad. Sci. USA*, 109(39), 15553–15559. https://doi.org/10.1073/pnas.1213423109.

Husson, F., Lê, S. and Pagès, J. (2017). Exploratory multivariate analysis by example using R: Chapman and Hall/CRC.

Johannes, F., Porcher, E., Teixeira, F.K. et al. (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* 5 (6), e1000530. https://doi.org/10.1371/journal.pgen.1000530.

Khan, M.A. and Korban, S.S. (2012) Association mapping in forest trees and fruit crops. *J. Exp. Bot.* 63(11), 4045–4060. https://doi.org/10.1093/jxb/ers105.

Klein, M. and Papenbrock, J. (2004) The multi-protein family of Arabidopsis sulphotransferases and their relatives in other plant species. *J. Exp. Bot.* 55(404), 1809–1820. https://doi.org/10.1093/jxb/erh183.

Kopriva, S., Hartmann, T., Massaro, G., Hönicke, P. and Rennenberg, H. (2004) Regulation of sulfate assimilation by nitrogen and sulfur nutrition in poplar trees. *Trees*, 18(3), 320–326. https://doi.org/10.1007/s00468-003-0309-4.

Lairson, L., Henrissat, B., Davies, G. and Withers, S. (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* 77, 521–555.

Lamara, M., Raherison, E., Lenz, P., Beaulieu, J., Bousquet, J. and MacKay, J. (2016) Genetic architecture of wood properties based on association analysis and co-expression networks in white spruce. *New Phytol.* 210 (1), 240–255. https://doi.org/10.1111/nph.13762.

Larocque, G.R. and Marshall, P.L. (1995) Wood relative density development in red pine (Pinus resinosa Ait.) stands as affected by different initial spacings. *For. Sci.* 41(4), 709–728.

Larsson, H., Källman, T., Gyllenstrand, N. and Lascoux, M. (2013) Distribution of Long-Range Linkage Disequilibrium and Tajima's D Values in Scandinavian Populations of Norway Spruce (Picea abies). *G3: Genes - Genomes - Genetics*, 3(5), 795–806. https://doi.org/10.1534/g3.112.005462.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997.*

Li, Z. and Sillanpää, M.J. (2013) A Bayesian nonparametric approach for mapping dynamic quantitative traits. *Genetics*, 194(4), 997–1016.

Li, Z. and Sillanpää, M.J. (2015) Dynamic quantitative trait locus analysis of plant phenomic data. *Trends Plant Sci.* 20(12), 822–833. https://doi.org/10.1016/j.tplants.2015.08.012.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, Z., Hallingbäck, H.R., Abrahamsson, S., Fries, A., Gull, B.A., Sillanpää, M.J. and García-Gil, M.R. (2014) Functional multi-locus QTL mapping of temporal trends in Scots pine wood traits. *G3: Genes - Genomes - Genetics*, **4**(12), 2365–2379. https://doi.org/10.1534/g3.114.014068.

Liu, H.-J. and Yan, J. (2019) Crop genome-wide association study: a harvest of biological relevance. *Plant J.* **97**(1), 8–18. https://doi.org/10.1111/tpj.14139.

Lu, M., Krutovsky, K.V., Nelson, C.D., Koralewski, T.E., Byram, T.D. and Loopstra, C.A. (2016) Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.). *BMC Genom.* **17**(1), 730.

Lundqvist, S.-O., Seifert, S., Grahn, T., Olsson, L., García-Gil, M.R., Karlsson, B. and Seifert, T. (2018) Age and weather effects on between and within ring variations of number, width and coarseness of tracheids and radial growth of young Norway spruce. *Eur. J. Forest Res.* **137**(5), 719–743.

Ma, C.-X., Casella, G. and Wu, R. (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics*, **161**(4), 1751–1762.

Mackay, T.F.C. (2013) Epistasis and quantitative traits: using model organisms to study gene–gene interactions. *Nat. Rev. Genet.* **15**, 22. https://doi.org/10.1038/nrg3627.

Maloney, V.J., Samuels, A.L. and Mansfield, S.D. (2012) The endo-1, 4-β-glucanase Korrigan exhibits functional conservation between gymnosperms and angiosperms and is required for proper cell wall formation in gymnosperms. *New Phytol.* **193**(4), 1076–1087.

Mauriat, M. and Moritz, T. (2009) Analyses of GA20ox- and GID1-over-expressing aspen suggest that gibberellins play two distinct roles in wood formation. *Plant J.* **58**(6), 989–1003.

McKown, A.D., Klápště, J., Guy, R.D., Geraldes, A., Porth, I., Hannemann, J., Friedmann, M., Muchero, W., Tuskan, G.A. and Ehlting, J. (2014) Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa. New Phytol.* **203**(2), 535–553. https://doi.org/10.1111/nph.12815.

Meinshausen, N. and Bühlmann, P. (2010) Stability selection. *J. R. Stat. Soc. Series B Stat. Methodol.* **72**(4), 417–473. https://doi.org/10.1111/j.1467-9868.2010.00740.x.

Micheli, F. (2001) Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends Plant Sci.* **6**(9), 414–419. https://doi.org/10.1016/S1360-1385(01)02045-3.

Namroud, M.-C., Guillet-Claude, C., Mackay, J., Isabel, N. and Bousquet, J. (2010) Molecular evolution of regulatory genes in spruces from different species and continents: heterogeneous patterns of linkage disequilibrium and selection but correlated recent demographic changes. *J. Mol. Evol.* **70**(4), 371–386.

Neale, D.B. and Savolainen, O. (2004) Association genetics of complex traits in conifers. *Trends Plant Sci.* **9**(7), 325–330.

Novaes, E., Osorio, L., Drost, D.R. *et al.* (2009) Quantitative genetic analysis of biomass and wood chemistry of *Populus* under different nitrogen levels. *New Phytol.* **182**(4), 878–890. https://doi.org/10.1111/j.1469-8137.2009.02785.x.

Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S. and Alexeyenko, A. (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature*, **497**(7451), 579–584.

Olesen, P. (1977) The variation of the basic density level and tracheid width within the juvenile and mature wood of Norway spruce. *For. Tree Improv.* **12**, 1–22.

Paine, C.E.T., Marthews, T.R., Vogt, D.R., Purves, D., Rees, M., Hector, A. and Turnbull, L.A. (2012) How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods Ecol. Evol.* **3**(2), 245–256. https://doi.org/doi:10.1111/j.2041-210X.2011.00155.x.

Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W. and Buerkle, C. (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol.* **21**(12), 2991–3005. https://doi.org/10.1111/j.1365-294X.2012.05513.x.

Peltola, H., Gort, J., Pulkkinen, P., Gerendiain, A.Z., Karppinen, J. and Ikonen, V.-P. (2009) Differences in growth and wood density traits in Scots

Pine (*Pinus sylvestris* L.) genetic entries grown at different spacing and sites. *Silva Fenn.* **43**(3), 339–354.

Phan, T.D., Bo, W., West, G., Lycett, G.W. and Tucker, G.A. (2007) Silencing of the major salt-dependent isoform of Pectinesterase in tomato alters fruit softening. *Plant Physiol.* **144**(4), 1960–1967. https://doi.org/10.1104/pp.107.096347.

Porth, I., Klápště, J., Skyba, O. *et al.* (2013) Genome-wide association mapping for wood characteristics in *Populus* identifies an array of candidate single nucleotide polymorphisms. *New Phytol.* **200**(3), 710–726. https://doi.org/doi:10.1111/nph.12422.

Resende, R.T., Resende, M.D.V., Silva, F.F., Azevedo, C.F., Takahashi, E.K., Silva-Junior, O.B. and Grattapaglia, D. (2017a) Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in Eucalyptus. *New Phytol.* **213**(3), 1287–1300. https://doi.org/10.1111/nph.14266.

Resende, R.T., Resende, M.D.V., Silva, F.F., Azevedo, C.F., Takahashi, E.K., Silva-Junior, O.B. and Grattapaglia, D. (2017b) Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in Eucalyptus. *New Phytol.* **213**(3), 1287–1300.

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003) *Semiparametric Regression* Vol. **12**. Cambridge University Press: UK.

Sewell, M.M., Bassoni, D.L., Megraw, R.A., Wheeler, N.C. and Neale, D.B. (2000) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). I. Physical wood properties. *Theor. Appl. Genet.* **101**(8), 1273–1281. https://doi.org/10.1007/s001220051607.

Shiu, S.-H. and Bleecker, A.B. (2001) Plant receptor-like kinase gene family: diversity, function, and signaling. *Sci. STKE*, **2001**(113), re22.

Sillanpää, M., Kontunen-Soppela, S., Luomala, E.-M., Sutinen, S., Kangasjärvi, J., Häggman, H. and Vapaavuori, E. (2005) Expression of senescence-associated genes in the leaves of silver birch (*Betula pendula*). *Tree Physiol.* **25**(9), 1161–1172.

Storey, J.D., Akey, J.M. and Kruglyak, L. (2005) Multiple locus linkage analysis of genomewide expression in yeast. *PLoS Biol.* **3**(8), e267. https://doi.org/10.1371/journal.pbio.0030267.

Strauss, S., Lande, R. and Namkoong, G. (1992) Limitations of molecular-marker-aided selection in forest tree breeding. *Can. J. For. Res.* **22**(7), 1050–1061.

Team, R. (2015). RStudio: integrated development for R. RStudio, Inc., Boston, MA *URL* http://www.rstudio.com.

Thavamanikumar, S., Southerton, S.G., Bossinger, G. and Thumma, B.R. (2013) Dissection of complex traits in forest trees—opportunities for marker-assisted selection. *Tree Genet. Genomes*, **9**(3), 627–639.

Thumma, B.R., Southerton, S.G., Bell, J.C., Owen, J.V., Henery, M.L. and Moran, G.F. (2010) Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens. Tree Genet. Genomes*, **6**(2), 305–317.

Tibshirani, R. (1996) Regression shrinkage and selection via the LASSO. *J. R. Stat. Soc. Series B Methodol.* **58**, 267–288.

Trujillo-Moya, C., George, J., Fluch, S., Geburek, T., Grabner, M., Karanitsch-Ackerl, S., Konrad, H., Mayer, K., Sehr, E. and Wischnitzki, E. (2018) Drought sensitivity of Norway Spruce at the Species' warmest fringe: quantitative and molecular analysis reveals high genetic variation among and within provenances. *G3 (Bethesda, Md.)*, **8**(4), 1225–1245.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D. and Thibault, J. (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*, **43**(1), 11.10.1-11.10.33.

Verhaegen, D., Plomion, C., Gion, J.-M., Poitel, M., Costa, P. and Kremer, A. (1997) Quantitative trait dissection analysis in Eucalyptus using RAPD markers: 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theor. Appl. Genet.* **95**(4), 597–608. https://doi.org/10.1007/s001220050601.

Vidalis, A., Scofield, D.G., Neves, L.G., Bernhardsson, C., García-Gil, M.R. and Ingvarsson, P. (2018) Design and evaluation of a large sequence-capture probe set and associated SNPs for diploid and haploid samples of Norway spruce (*Picea abies*). *bioRxiv*. https://doi.org/10.1101/291716.

Wegrzyn, J.L., Eckert, A.J., Choi, M., Lee, J.M., Stanton, B.J., Sykes, R., Davis, M.F., Tsai, C.-J. and Neale, D.B. (2010) Association genetics of

traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. *New Phytol.* **188**(2), 515–532. https://doi.org/doi:10.1111/j.1469-8137.2010.03415.x.

Wu, W.-R., Li, W.-M., Tang, D.-Z., Lu, H.-R. and Worland, A. (1999) Time-related mapping of quantitative trait loci underlying tiller number in rice. *Genetics*, **151**(1), 297–303.

Wu, R., Ma, C.-X., Lin, M. and Casella, G. (2004) A general framework for analyzing the genetic architecture of developmental characteristics. *Genetics*, **166**(3), 1541–1551.

Xing, J.U.N., Li, J., Yang, R., Zhou, X. and Xu, S. (2012) Bayesian B-spline mapping for dynamic quantitative traits. *Genet. Res.* **94**(2), 85–95. https://doi.org/10.1017/S0016672312000249.

Yamasaki, K., Kigawa, T., Inoue, M. *et al.* (2004) Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1. *Plant Cell*, **16**(12), 3448–3459.

Yang, R. and Xu, S. (2007) Bayesian shrinkage analysis of quantitative trait loci for dynamic traits. *Genetics*, **176**(2), 1169–1185. https://doi.org/10.1534/genetics.106.064279.

Yang, R., Tian, Q. and Xu, S. (2006) Mapping quantitative trait loci for longitudinal traits in line crosses. *Genetics*, **173**(4), 2339–2356.

Yang, J., Li, L., Jiang, H., Nettleton, D. and Schnable, P.S. (2014) Dominant gene action accounts for much of the missing heritability in a gwas and provides insight into heterosis. *Genome-wide association studies to dissect the genetic architecture of yield-related traits in maize and the genetic basis of heterosis*, **1001**, 44.

Zhong, R., Burk, D.H., Morrison, W.H. and Ye, Z.-H. (2002) A kinesin-like protein is essential for oriented deposition of cellulose microfibrils and cell wall strength. *Plant Cell*, **14**(12), 3101–3117.

Zhou, G., Chen, Y., Yao, W., Zhang, C., Xie, W., Hua, J., Xing, Y., Xiao, J. and Zhang, Q. (2012) Genetic composition of yield heterosis in an elite rice hybrid. *Proc. Natl Acad. Sci. USA*, **109**(39), 15847–15852. https://doi.org/10.1073/pnas.1214141109.

Zou, H. (2006) The adaptive LASSO and its oracle properties. *J. Am. Stat. Assoc.* **101**(476), 1418–1429.

III

ORIGINAL ARTICLE

Plant, Cell & Environment   WILEY

Association genetics identifies a specifically regulated Norway spruce laccase gene, *PaLAC5*, linked to *Heterobasidion parviporum* resistance

Malin Elfstrand[1] | John Baison[2] | Karl Lundén[1] | Linghua Zhou[2] |
Ingrid Vos[3] | Hernan Dario Capador[1] | Matilda Stein Åslund[1] | Zhiqiang Chen[2] |
Rajiv Chaudhary[1] | Åke Olson[1] | Harry X. Wu[2] | Bo Karlsson[3] | Jan Stenlid[1] |
María Rosario García-Gil[2]

[1]Uppsala Biocentre, Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden

[2]Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, Umeå, Sweden

[3]Skogforsk, Svalöv, Sweden

Correspondence
Malin Elfstrand, Uppsala Biocentre, Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, P.O.Box 7026 75005 Uppsala, Sweden.
Email: malin.elfstrand@slu.se

Funding information
Swedish Foundation for Strategic Research, Grant/Award Number: RBP14-0040; The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning, Grant/Award Number: 2017-0040; VINNOVA, Grant/Award Numbers: 2015-02290, 2016-00504

Abstract

It is important to improve the understanding of the interactions between the trees and pathogens and integrate this knowledge about disease resistance into tree breeding programs. The conifer Norway spruce (*Picea abies*) is an important species for the forest industry in Europe. Its major pathogen is *Heterobasidion parviporum*, causing stem and root rot.

In this study, we identified 11 Norway spruce QTLs (Quantitative trait loci) that correlate with variation in resistance to *H. parviporum* in a population of 466 trees by association genetics. Individual QTLs explained between 2.1 and 5.2% of the phenotypic variance. The expression of candidate genes associated with the QTLs was analysed in silico and in response to *H. parviporum* hypothesizing that (a) candidate genes linked to control of fungal sapwood growth are more commonly expressed in sapwood, and; (b) candidate genes associated with induced defences are respond to *H. parviporum* inoculation. The Norway spruce *laccase PaLAC5* associated with control of lesion length development is likely to be involved in the induced defences. Expression analyses showed that *PaLAC5* responds specifically and strongly in close proximity to the *H. parviporum* inoculation. Thus, *PaLAC5* may be associated with the lignosuberized boundary zone formation in bark adjacent to the inoculation site.

KEYWORDS

genome-wide association study (GWAS), lignosuberized boundary zone, mitochondrion, sapwood, secretory and endosomal trafficking pathways, suberin, TOM40

1 | INTRODUCTION

The importance of trees and forests for sustaining terrestrial life and biodiversity can probably not be exaggerated (Petit & Hampe, 2006).

Pathogen and pest attacks on trees negatively impact the health and biodiversity of native forest ecosystems as well as forest plantations, which can have large economic, ecological and societal consequences (Cubbage, Pye, Holmes, & Wagner, 2000; Garbelotto & Gonthier, 2013; Pautasso, Schlegel, & Holdenrieder, 2015; Woodward, Stenlid, Karjalainen, & Hüttermann, 1998). Therefore, it is important to increase the understanding of interactions between the tree and a pathogen in order to incorporate traits that confer to increased resistance into forest tree breeding programs.

Norway spruce [*Picea abies* (L.) Karst.] is economically important for the forest industry in Europe. Its major pathogens are fungi in the species complex *Heterobasidion annosum* sensu lato (s.l.), which causes stem and root rot in Norway spruce and several other conifer tree species (Garbelotto & Gonthier, 2013; Woodward et al., 1998). Under natural conditions, airborne spores of *H. annosum* s.l. can infect stumps created after harvesting and thinning operations. Once the stump is infected, surrounding trees or stumps can be infected by secondary spread when *H. annosum* s.l. mycelium enters neighbouring trees through root grafts and contacts (Oliva, Bendz-Hellgren, & Stenlid, 2011; Redfern & Stenlid, 1998). In Norway spruce, resistance to the spruce-infecting congener *Heterobasidion parviporum* is quantitative in its nature (Arnerup, Swedjemark, Elfstrand, Karlsson, & Stenlid, 2010; Chen et al., 2018; Karlsson & Swedjemark, 2006; Steffenrem, Solheim, & Skrøppa, 2016), and classical interval mapping-based quantitative trait locus (QTL) analysis for resistance to *H. parviporum* identified 13 QTL linked to host resistance (Lind et al., 2014). *PaLAR3*, on the QTLs associated with control of fungal spread in the sapwood, has been validated and the function of the variation at the locus described (Nemesio-Gorriz et al., 2016).

A feature that Norway spruce has in common with all tree species is that a large fraction of the biomass is invested in the sapwood in the trunk (Petit & Hampe, 2006). The primary function of the sapwood is to transport water and nutrients to the crown and it is dominated by dead cells that have a limited capacity to respond to biotic or abiotic stress (Johansson & Theander, 1974; Oliva et al., 2015; Shain, 1971). To protect the sapwood, the trunk of a tree is clad in an impermeable barrier, bark. The term "bark" commonly refers to all tissues external to the vascular cambium of trees. The outer bark is highly suberized and lignified, making it extremely resistant to mechanical and chemical degradation. Only a few pathogenic microorganisms are capable of directly penetrating the outer bark (Lindberg & Johansson, 1991). Therefore, a common mode of entry for fungi that cause stem cankers and decays is via mechanical wounds, exposing the cortex, secondary phloem tissues or the xylem (Woodward & Pocock, 1996). The speed at which the tree is able to seal off the tissues exposed by wounding with wound periderm is critical in avoiding damaging infections and subsequent loss of water transport capacity. The process to heal the bark begins with rapid necrosis of cells closest to the wound or progressing infection. It then continues with programmed death of cells adjacent to the necrosis, forming the lignosuberized boundary zone (LSZ), and de-differentiation of cells next to the LSZ followed by differentiation of the wound periderm

(Bodles, Beckett, & Woodward, 2007; Mullick, 1977; Woodward, Bianchi, Bodles, Beckett, & Michelozzi, 2007).

The trait control of lesion length extension (LL, with reported heritability values of 0.14–0.33) is measured as the size of the discernible necrosis cells closest to the wound or progressing infection (Arnerup, Lind, Olson, Stenlid, & Elfstrand, 2011; Chen et al., 2018; Steffenrem et al., 2016). It could be argued that LL provides a measure of how the induced defences and wound healing responses interact to control the spread of the necrotrophic pathogen (Arnerup et al., 2011; Chen et al., 2018; Danielsson et al., 2011; Lind et al., 2014; Steffenrem et al., 2016). The trait control of fungal spread in the sapwood (fungal sapwood growth, SWG) can be considered to provide a measure of how well the combination of constitutive defences and the induced defence responses in the parenchymatic cells can control the spread of *H. parviporum* in the exposed sapwood (Johansson & Stenlid, 1985; Oliva et al., 2015). The narrow-sense heritability of SWG has been estimated to vary between 0.11 and 0.42 depending on the material studied (e.g., experimental cross, natural population) (Arnerup et al., 2010; Chen et al., 2018).

To date, the main focus of practical breeding in Norway spruce has been on climatic adaptation, growth and wood quality traits (Skrøppa, Solheim, & Steffenrem, 2015). In contrast, breeding for replantation material with improved resistance to *H.annosum* s.s. and *H. parviporum* is an overlooked objective because of limited information about genetic variation in resistance to these pathogens and the lack of reliable selection techniques (Skrøppa et al., 2015). There are, however, clearly sufficient phenotypic and genetic variation for resistance to *H. parviporum* in Norway spruce to allow for breeding (Arnerup et al., 2010; Chen et al., 2018; Karlsson & Swedjemark, 2006; Steffenrem et al., 2016), and no adverse correlations between resistance to *H. parviporum* and growth or wood properties traits (Chen et al., 2018; Steffenrem et al., 2016). Hence, the selection for *H. parviporum* resistance in breeding programmes could lead to considerable gain without compromising other breeding achievements (Chen et al., 2018).

To gain a deeper understanding of the heritability and genetic architecture of, for example, disease resistance traits, including the number, location, effect and nature of the loci involved, quantitative and molecular genetic approaches can be used to analyse the relationships between DNA polymorphism and phenotypic variation (Bartholomé et al., 2016; Neale & Savolainen, 2004). The two main approaches to detect QTLs: Interval mapping (IM) in experimental crosses or linkage disequilibrium (LD) mapping, commonly known as genome-wide association studies (GWAS) (Neale & Savolainen, 2004). GWAS, relying on historical recombination in the mapping population, overcomes the limited resolution of IM in experimental crosses (Baison et al., 2019; Neale & Savolainen, 2004). If enough markers can be analysed, this should be especially advantageous in conifers that have particularly short average distances of maintained LD, often even confined within genes (Namroud, Guillet-Claude, Mackay, Isabel, & Bousquet, 2010). The effects of LD are also influenced by the extreme physical distances separating genes in conifers (Nystedt et al., 2013).

It is likely that the Norway spruce genome harbours additional, yet undetected loci, to the 13 QTLs already identified by (Lind et al., 2014) controlling resistance to *H. parviporum* (Chen et al., 2018; Hall, Hallingbäck, & Wu, 2016). Identification of further loci would support the initiation of a breeding programme for the resistance to the pathogen in Norway spruce and, just as importantly, improve the understanding of the interactions between trees and necrotrophic pathogens. The short maintained LD and the polygenic nature of the traits controlling resistance suggest that GWAS could be a powerful method to identify further QTL regions associated with *H. parviporum* resistance in Norway spruce. Consequently, in this study, we aimed to identify Norway spruce loci that correlate with variation in resistance to *H. parviporum* in a population of 466 Norway spruce trees by GWAS. We identified candidate genes associated with the QTLs and analysed the expression patterns of the candidate genes in response to *H. parviporum* hypothesizing that (a) candidate genes linked to the SWG trait would be expressed in sapwood while candidate genes linked to LL are expressed in more peripheral tissues, and; (b) candidate genes that are part of the induced defence are induced in response to *H. parviporum* inoculation.

2 | MATERIALS AND METHODS

2.1 | Phenotyping of resistance traits in the progeny of 466 Norway spruce mother trees

We used the currently available largest Norway spruce resistance phenotyping dataset to perform the GWAS. The material, inoculation method and genetic analyses are described in detail in (Chen et al., 2018). On average ten 2-year-old, open-pollinated progenies derived from 466 tested plus trees in the Swedish breeding population were inoculated with *H. parviporum* Niemelä & Korhonen strain Rb175. A wooden dowel colonized by *H. parviporum* was fixated at a wound on the stem of the plant with Parafilm. The inoculated plants were kept under ambient light and temperature in the forest tree nursery and harvested 21 days post-inoculation. The induced defence responses (LL) in the phloem and inner bark were estimated by measuring the discernible lesion spread upwards and downwards from the edge of the inoculation point on the inside of the bark. SWG was estimated using established protocols (Arnerup et al., 2010; Stenlid & Swedjemark, 1988) (Table 1). The seedlings were cut up into five mm discs and placed on moist filter papers in Petri dishes. Plates were incubated in darkness under moist conditions at 21°C for 1 week to induce conidia formation. Thereafter, the presence or absence of *H. parviporum* conidia on each individual disc was determined under a stereomicroscope. For each seedling, the sum of the discs where conidia were observed multiplied by 5 (mm) was noted as SWG. Plates where no conidia could be observed on the discs, the inoculation point and on the inoculation plug, and that showed total lesion length of 2 mm or shorter, were treated as inoculation failures and were discarded (Lind et al., 2014). Chen et al. (2018) reported narrow-sense heritability values of 0.33 and 0.42, respectively, for LL and SWG and

TABLE 1 Summary statistics of the phenotype data used in the trait-marker association study (Details can be found in Chen et al. (2018))

Inoculation study	Acron.	Unit	Na	Mean
Diameter[b]	D	mm	4,628	4.0
Lesion length[c]	LL	mm	4,547	7.6
Fungal growth[d]	FG/SWG	mm	4,554	32.5
Vitality[e]	Vitality	Classes	4,376	1.9

[a]N: total number of progenies with valid recording of the trait.
[b]Diameter of the progenies at the inoculation site.
[c]Length of the necrotic lesion in the phloem and inner bark.
[d]Fungal growth in the sapwood of the progenies.
[e]Vitality of the progenies where score 1 was given to fully vital and worst score 3 was given to plants showing a pronounced loss of vitality.

moderate phenotypic (0.48) and genetic (0.47) correlations between LL and SWG in this material.

2.2 | Norway spruce genotyping and SNP annotation

Dormant buds were collected from each of the mother trees. Total genomic DNA was extracted from the buds, using the Qiagen Plant DNA extraction kit (Qiagen, Hilden, Germany), and the DNA was quantified using the Qubit® ds DNA Broad Range (BR) Assay Kit (Oregon, USA). The generation and evaluation of exome capture for Norway spruce are described elsewhere (Vidalis et al., 2018). Sequence capture on the mother tree DNA was performed using 40,018 previously evaluated diploid probes (Baison et al., 2019; Vidalis et al., 2018). Probe design and sequence capture were done by RAPiD Genomics (Gainesville, FL, USA). In brief, Illumina sequencing compatible libraries were amplified with 14 cycles of PCR and the probes were then hybridized to a pool comprising 500 ng of eight equimolarly combined libraries following Agilent's SureSelect Target Enrichment System (Agilent Technologies). These enriched libraries were then sequenced to an average depth of 15x using an Illumina HiSeq 2,500 (San Diego, USA) on the 2 × 100 bp sequencing mode.

Read mapping and initial variant calling as well as the recalibration of the quality of SNP calling were then applied to filter the raw variants, described in detail in Baison et al. (2019). In brief, the variant calling was made using GATK HaplotypeCaller v.3.6 as per the best practices protocol (Auwera et al., 2013) in gVCF output format. To increase accuracy, hard filters in the form of minor allele frequency (MAF) and "missingness" of <0.05 and >20%, respectively, were then performed on the final dataset.

3 | GWAS

The LASSO model as described by Li et al. (2014) was applied to the *H. parviporum* resistance trait data for the detection of QTLs.

The LASSO model:

$$\min_{(\alpha_0, \alpha_j)} \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \alpha_o - \sum_{j=1}^{P} x_\gamma - \alpha_j \right)^2 + \lambda \sum_{j=1}^{P} \alpha_j, \qquad (1)$$

where y_i is the estimated breeding values (EBV) of an individual i ($i = 1,...,n$; n is the total number of individuals) for each trait, α_0 is the population mean parameter, x_{ij} is the genotypic value of individual i and marker j coded as 0, 1 and 2 for three marker genotypes AA, AB and BB, respectively, α_j is the effect of marker j ($i = 1,...,n$; n is the total number of markers) and λ (>0) is a shrinkage tuning parameter. A fundamental idea of LASSO is to utilize the penalty function to shrink the SNP effects towards zero, and only keep a small number of important SNPs that are highly associated with the trait in the model. The stability selection probability (SSP) of each SNP being selected to the model was applied as a way to control the false discovery rate and determine significant SNPs (H. Gao et al., 2014; Li & Sillanpää, 2015). For a marker to be declared significant, an SSP inclusion ratio (Frequency) was used with an inclusion frequency of all traits. This frequency inferred that the expected number of falsely selected markers was less than one, according to the formula of Bühlmann, Kalisch, and Meier (2014). Population structure was accounted for in all analyses by including principal components based on the genotype data as covariates into the model (Baison et al., 2019). An adaptive LASSO approach (Baison et al., 2019; Zou, 2006) was used to determine the percentage of phenotypic variance (PVE) (H^2_{QT}) of all the QTLs. The analyses were all performed in RStudio (Team, 2015).

3.1 | Identification of candidate genes associated with the QTLs

To assess putative functionality of SNPs with significant associations, a gene enrichment analysis of putative genes and their associated orthologs was performed against the *P. abies* v1.0 genome (http://congenie.org), collecting PFAM and GO term annotations and *Populus* and *Arabidopsis* orthologues. The position of the detected QTLs in Norway spruce genome was estimated by searching an ultra-dense genetic map (Bernhardsson et al., 2019) for markers derived from the same probes as the SNP markers holding the QTLs, identified based on tblastn sequence homology for the SNP array sequences in the Lind et al. (2014) study, as described by (Bernhardsson et al., 2019).

Information on the expression pattern of the putative candidate genes associated with the QTL, in the Norway spruce clone Z4006 (the clone sequenced in Nystedt et al. (2013)) and in wood, were collected from three sources. Firstly, expression data were downloaded from the publicly available *P.abies* exAtlas (https://www.congenie.org) and NorWood v1.0 (http://norwood.congenie.org) databases, respectively. Both these databases are comprised of expression profiles from approximately 50-year-old ramets of the genotype "Z4006." Then, we examined an RNAseq study of bark and phloem samples harvested at seven dpi proximal (0–5 mm from the wound) and distal to the

inoculation site (10–15 mm away from the wound) from two Norway spruce genotypes (S21K0220126 and S21K0220184) inoculated with *H. parviporum* (Chaudhary et al., submitted manuscript). In brief, two-year-old branches on clones of S21K0220126 and S21K0220184 were inoculated and sampled as described above using wounding as a control. A total RNA from three biological replicates of each clone per treatment were sequenced on the Illumina HiSeq 2500 at the SNP&SEQ Technology Platform (SciLifeLab, Uppsala). Quality filtering was done using Nesoni 0.97 (http://www.vicbioinformatics.com/nesoni-cookbook/index.html#). Differential gene expression was identified using the Tophat-cufflinks pipeline (Trapnell et al., 2012, 2014; Trapnell et al., 2013) and the "*P. abies* v1.0-all-cds.fna" gene catalogue as a reference (Chaudhary et al., submitted manuscript).

3.2 | Branch inoculation with *H. parviporum*

We performed an inoculation experiment on six-year-old grafted cuttings of the Norway spruce genotype S21K7820222. Branches on healthy-looking potted plants were inoculated with wooden dowels colonized by *H. parviporum* Rb175 fixated to a wound on a two-year-old branch with Parafilm. Control treatment branches were wounded and covered with Parafilm. The inoculated plants kept at ambient light and temperature conditions in a greenhouse. At 7 days post-inoculation (dpi) bark surrounding the wounds and inoculation sites were cut into two sections and samples were collected at the inoculation site 0–5 mm around the wound and distal to the inoculation site 10–15 mm from the wound. The bark samples were frozen separately in liquid nitrogen and stored at −80°C until further use.

3.3 | Quantitative PCR analysis of expression patterns in response to H. parviporum inoculation

The total RNA was isolated according to the protocol by Chang, Puryear, and Cairney (1993). To eliminate genomic DNA contamination, samples were treated with DNase I (Sigma-Aldrich) according to the manufacturer's instructions. RNA integrity and quantity were analysed by using the Agilent RNA 6000 Nano kit (Agilent Technologies Inc.). The 1 μg of total RNA was reverse transcribed to cDNA with the iScript cDNA Synthesis Kit (Bio-Rad) in a total reaction volume of 20 μl according to the manufacturer's instructions, followed by a two-fold dilution of the cDNA and storage at − 20°C.

Quantitative PCR (qPCR) reactions were performed with the SsoFast™ EvaGreen® Supermix (Bio-Rad) according to the instructions in the manual, using 0.3 μM of each primer (Table S1 in Data S1) and Norway spruce cDNA equivalent to 25 ng of total RNA. The qPCRs were carried out in an iQ5™ Multicolor Real-Time PCR Detection System thermocycler (Bio-Rad) using a program with a 30 s initial denaturation step at 95°C, followed by 40 cycles of 5 s denaturation at 95°C and 10 s at 60°C. Melt curve analyses were used to validate the amplicon. Four biological replicates were used per treatment and two

technical repetitions per standard, sample and negative control were run.

The relative expression was calculated from threshold cycle (Ct) values using the 2ΔΔCT-method (Livak & Schmittgen, 2001) by using the geometric mean of *Phosphoglucomutase* (Vestman et al., 2011) and *elongation factor 1-α* (ELF1α) (Arnerup et al., 2011) to normalize transcript abundance. The gene expression experiments were performed with four biological and two technical replicates. One-way ANOVA with Dunns Post-test (GraphPad Prism 5.0) was used to detect differences in expression levels between treatments.

4 | RESULTS

4.1 | Trait association mapping identifies novel QTLs for resistance to *H. parviporum*

From an average of 1.5 million paired end sequence reads per individual, 197,399 high confidence SNPs from 23,837 probes were identified. The majority of the SNPs were missense (61%) and silent (36%), the highest percentage being either upstream or downstream variants (68% total).

Employing a Stability Selection Probability (SSP) on the estimated breeding values (EBVs) for SWG and LL of the offspring on the 466 trees, we identified six SNPs with significant associations for SWG and five SNPs associating with LL (Table 2). The QTLs for control of sapwood growth of *H. parviporum* (SWG) explained similar fractions of the observed phenotypic variation (H^2_{QTL}) 2.4 to 5.2% (Table 2). The five QTLs for control of the LL development in bark explained between 2.1 and 4.4% of the observed phenotypic variation (Table 2).

To investigate if the identified QTLs are independent from previously identified QTLs for resistance to the same isolate of *H. parviporum* using IM (Lind et al., 2014), we searched an ultra-dense genetic map (Bernhardsson et al., 2019) for the probes the SNP markers originated from. This allowed us to estimate the position of the detected QTLs and the original IM-based QTLs in the Norway spruce genome. We could estimate the position in the Norway spruce genome for six of the SNPs/probes (Table S2.I and Figure S2.II in Data S1). All of the identified SNPs/probes were positioned >30 cM away from the original IM-based QTLs in the genetic map. Given that the maintained LD is estimated to only 109 bp across all the tagged genomic sequences in this study (Table S2 in Data S1), it is likely that they are independent. The SNP MA_53835_9763, associating with the trait SWG, presented a potential exception as the probe MA_14663 is positioned 4 cM away from MA_53835 in the map (Bernhardsson et al., 2019). The probe MA_14663 corresponds to the SNP array sequence for an IM-based QTL for infection prevention (Lind et al., 2014; Chaudhary et al., submitted manuscript).

On the scaffolds holding the SNPs associated with the resistance traits, a total of 14 gene models were identified, including 11 high- or medium-quality Norway spruce gene models (Table 3). On the scaffolds holding more than one gene model, the SNPs were positioned in MA_5978g0020, MA_25569g0020 and MA_97119g0010. Seven of the candidate genes associated with SWG QTLs and seven with LL (Tables 2 and 3). PFAM and GO term annotations and *Populus* and *Arabidopsis* orthologues were collected from *P. abies* v1.0 genome portal (Table 3). These metrics suggested that the gene models MA_97119g0010 and MA_97119g0020, found on the scaffold harbouring the SNP MA_97119_12277, indeed represented one gene. BlastN searches against the NCBI database essentially confirmed this suggestion as both gene models match JX500691.1 (*Picea abies*

TABLE 2 Significant association in the GWA study

Phenotype[a]	QTL	SNP[b]	Allele[c]	SNP feature[d]	Frequency[e]	PVE (%)[f]
SWG_tot	8675	MA_5978_21,011	T/C	Missense	0.71	4.83
	26756	MA_17884_58584	A/G	Upstream variant	0.72	3.41
	54184	MA_53072_3732	G/A	Synonymous	0.551	2.88
	54695	MA_53835_9763	G/A	Upstream variant	0.567	2.40
	56105	MA_56128_7752	C/A	Upstream variant	0.545	5.21
	71928	MA_84091_11329	C/A	Upstream variant	0.534	2.23
LL_tot	21105	MA_14352_27165	G/A	Missense variant	0.603	3.82
	27795	MA_18316_3165	G/T	Upstream variant	0.618	2.11
	31060	MA_19645_22184	C/T	Missense	0.682	2.73
	37057	MA_25569_28091	T/C	Upstream variant	0.667	2.77
	81488	MA_97119_12277	T/C	Upstream variant	0.742	4.39

[a]Phenotype specifies the trait upon which the marker associate.
[b]SNP: The SNP name was composed of the contig (MA_number) and SNP position on contig. For example, the first SNP MA_5978_21011 was located on contig MA_5978 at position 21011 bp.
[c]Allele indicates the biallelic SNP.
[d]SNP feature allelic variation associated with the SNP.
[e]Frequency, stability selection probability inclusion ratios for markers declared significant.
[f]PVE, phenotypic variance explained, only values larger than 1.0% are displayed.

TABLE 3 Candidate Norway spruce gene models associated with the QTL markers

SNP[a]	Candidate gene[b]	Description (Blast2Go)[c]	PFAM-Description/GO term[d]	Orthologs populus/ Arabidopsis[e]
MA_5978_21,011	MA_5978g0010	Phenylcoumaran benzylic ether reductase	PF00106-short chain dehydrogenase, PF01073-3-beta hydroxysteroid dehydrogenase/isomerase family PF01118-Semialdehyde dehydrogenase, NAD binding domain, PF01370-NAD-dependent epimerase/ dehydratase family, PF02719-Polysaccharide biosynthesis protein, PF03435-Saccharopine dehydrogenase, PF03807-NADP oxidoreductase coenzyme F420-dependent, PF05368-NmrA-like family, PF07993-Male sterility protein, PF08659-KR domain, PF13460-NADH(P)-binding	Potri.009G118100.1/ AT1G75280.1
	MA_5978g0020	Nuclear factor 1 A-type isoform 2	PF06219-Protein of unknown function (DUF1005)	Potri.013G071000.3/ AT5G17640.1
MA_17884_58584	MA_17884g0010	Mitochondrial import receptor subunit TOM40-1	PF01459-Eukaryotic porin	Potri.007G000200.1/ AT3G20000.1
MA_53072_3732	MA_53072g0010			
MA_53835_9763	MA_53835g0010	Probable tocopherol O-chloroplastic	PF01209-ubiE/COQ5 methyltransferase family, PF01728-FtsJ-like methyltransferase, PF02353-Mycolic acid cyclopropane synthetase, PF03059-Nicotianamine synthase protein, PF05175-Methyltransferase small domain, PF05891-AdoMet dependent proline di-methyltransferase, PF07021-Methionine biosynthesis protein MetW, PF08003-Protein of unknown function (DUF1698), PF08241-Methyltransferase domain, PF08242-Methyltransferase domain, PF12847-Methyltransferase domain, PF13489-Methyltransferase domain, PF13578-Methyltransferase domain, PF13649-Methyltransferase domain, PF13659-Methyltransferase domain, PF13679-Methyltransferase domain, PF13847-Methyltransferase domain	Potri.013G077000.1 AT1G64970.1
MA_56128_7752	MA_56128g0010			Potri.006G130600.1
MA_84091_11329	MA_84091g0010			
MA_14352_27165	MA_14352g0010	Transcription factor bHLH118	PF00010-Helix–loop–helix DNA-binding domain	Potri.015G134300.1/ AT4G25400.1

TABLE 3   (Continued)

SNP[a]	Candidate gene[b]	Description (Blast2Go)[c]	PFAM-Description/GO term[d]	Orthologs populus/ Arabidopsis[e]
MA_18316_3165	MA_18316g0010	IST1 homologue	PF03398-Regulator of Vps4 activity in the MVB pathway	Potri.019G087400.1/ AT1G34220.2
MA_19645_22184	MA_19645g0010			
MA_25569_28091	MA_25569g0010		*GO:0005618-cell wall,* *GO:0016020-membrane,* *GO:0044444-cytoplasmic part*	Potri.002G054900.1/ AT1G03230.1
	MA_25569g0020			Potri.001G266500.1
MA_97119_12277	MA_97119g0010	Laccase	PF07732-Multicopper oxidase	Potri.019G124300.1 / AT2G30210.1
	MA_97119g0020	Laccase 12	PF00394-Multicopper oxidase, PF07731-Multicopper oxidase	Potri.010G183500.1 / AT5G05390.1

[a]SNP: The SNP name was composed of the contig (MA_number) and SNP position on contig.
[b]Candidate gene.
[c]Description (Blast2Go).
[d]PFAM-Description or GO terms when PFAM descriptions were missing.
[e]*Populus/Arabidopsis* orthologs identified in the *P. abies* v1.0 genome portal.

laccase LAC5a) with E = 4*10^{-135} and E = 0 and 99.62 and 99.71% identity, respectively. This laccase, *PaLAC5*, was originally isolated from lignin-forming Norway spruce suspension cultures. Apart from MA_97119, two other QTL holding scaffolds (MA_5978 and MA_25569) harboured more than one gene model (Table 3). Both of these scaffolds appear to hold two different gene models as judged by the PFAM annotations and *Populus* or *Arabidopsis* orthologs (Table 3). MA_5978g0010 appears to encode a phenylcoumaran benzylic ether reductase (PCBER) with similarity to PicglPPR21 (Porth, Hamberger, White, & Ritland, 2011). The gene model MA_14352g0010 may belong to the *basic helix–loop–helix (bHLH) DNA-binding* superfamily since the PFAM-ID PF00010 (Helix–loop–helix DNA-binding domain) is associated with the gene model. The candidate gene MA_18316g0010 is associated with PF03398 (regulator of Vps4 activity in the MVB pathway), indicating that this gene too may be involved in regulatory activities. The gene model MA_53835g0010 appears to encode a protein with methyltransferase capacities based on its PFAM annotation (Table 3) and its *Arabidopsis* orthologue (Table 3, and based on its PFAM annotation (PF01459) and the annotation of the *Arabidopsis* orthologue, AT3G20000.1 (Table 3) which encodes β-barrel protein, TOM40, forming channels in the outer mitochondrial membranes, it is likely that the candidate gene MA_17884g0010 encodes a Norway spruce TOM40-like protein.

4.2 | A majority of the candidate genes associated with SWG are expressed in stem and wood forming tissues

To gain a better understanding of the functionality of the candidate genes, we assessed the expression in silico using available resources
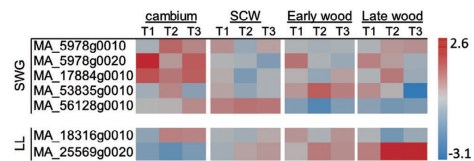
FIGURE 1   Relative expression levels of candidate genes associated to *H. parviporum* resistance QTLs through different stages of xylem development including cambium and expanding early wood (cambium), secondary cell wall-forming xylem (SCW), first dead early wood cells (Early wood) and the previous year's latewood (late wood). Data collected from NorWood v1.0 (http://norwood.congenie.org) database, T1-T3 represent the expression level in each of the three analysed trees (Jokipii-Lukkari et al., 2017). The bar to the left indicates the relative expression level of the candidate gene in the heat map [Colour figure can be viewed at wileyonlinelibrary.com]

such as NorWood and *P. abies* exATLAS databases. It predicted that the candidate genes linked to SWG would more commonly be expressed in sapwood than genes linked to LL. Only seven candidate genes (MA_5978g0010, MA_5978g0020, MA_17884g0010, MA_53835g0010, MA_56128g0010, MA_18316g0010 and MA_25569g0020) were expressed in any of the libraries in NorWood (Figure 1). Of the expressed candidate genes, five were linked to SWG. This indicated a trend (Chi-square = 3.233, *p* = .07) where candidate genes linked to the SWG QTLs were expressed more often in wood compared to candidate genes linked to LL.

NorWood is a database of transcript abundances in high spatial resolution section series throughout the cambial and woody tissues of Norway spruce (Jokipii-Lukkari et al., 2017). Three of the five candidate genes associated with control of SWG (MA_5978g0010, MA_5978g0020 and MA_17884g0010) showed the highest transcript

levels in the cambial region. MA_56128g0010, also associated with SWG, appeared to be more active in the expanding early wood and secondary cell wall-forming tissues (Figure 1). One of the two candidate genes associated with the LL extension in the phloem and inner bark that were detected in the NorWood libraries, MA_25569g0020 showed very high activity in the samples collected at the visual appearance of dead early wood cells and in latewood (Figure 1). The inspection of the expression patterns in the *P. abies* exATLAS indicated that all candidate genes but MA_84091g0010 and MA_19645g0010 were expressed in at least one tissue of the clone Z4006 (Figure S3 in Data S1). Apart from the candidate genes that were also detected in the NorWood database, several candidate genes (MA_14352g0010, MA_25569g0010, MA_97119g0010 and MA_97119g0020) associated with LL were found to be expressed in samples derived from stem tissues (Figure S3 in Data S1).

4.3 | The transcriptional responses to *H. parviporum* inoculation identifies candidates responding specifically to the pathogen

If the candidate gene models associated with QTLs contribute to the control of the *H. parviporum* infection, they may be involved in either the constitutive or induced defence in the tissue (or both) (Arnerup et al., 2011; Danielsson et al., 2011; Oliva et al., 2015). Assuming that genes associated with the induced defences respond to inoculation with the pathogen, it is relevant to assess the candidate genes expression pattern in response to *H. parviporum* (Arnerup et al., 2011; Danielsson et al., 2011; Oliva et al., 2015). We used an RNASeq study of transcriptional responses in bark and phloem response to wounding and *H. parviporum* inoculation (Chaudhary et al., submitted manuscript). Five candidate genes showed constitutive expression at seven dpi irrespective of the treatment: MA_5978g0020, MA_17884g0010, MA_53835g0010, MA_56128g0010

FIGURE 2 Expression profile of candidate genes for *H. parviporum* resistance in response to *H. parviporum* inoculation and wounding at seven dpi proximally (0–5 mm from the inoculation site) and distally (10–15 mm from the inoculation site) in the clones S21K0220126 and S21K0220184 (Chaudhary submitted MS). Asterisks indicate significant different expression levels between the inoculation treatment and the wounding control in *Cuffdiff*. The bar to the left indicates the FPKM values associated with the gene model [Colour figure can be viewed at wileyonlinelibrary.com]

and MA_25569g0020 (Figure 2). Most of these showed moderate expression levels, but MA_17884g0010 expression was relatively high in all samples. Four gene models associated with LL were differentially expressed at seven dpi: MA_14352g0010, MA_18316g0010, MA_97119g0010 and MA_97119g0020 (Figure 2). Interestingly, the two candidate gene models, (MA_97119g0010 and MA_97119g0020, i.e., *PaLAC5*) that showed the largest induction in response to the inoculation treatment compared to the wounding control proximal to the inoculation site, were not induced but rather downregulated distally at seven dpi (Figure 2). To validate the transcriptional responses estimated from the RNAseq data, we set up a separate inoculation experiment in a single Norway spruce genotype for qPCR validation of the expression patterns at seven dpi. The qPCR verified the transcriptional regulation patterns between *H. parviporum* inoculation and wounding treatment for most genes (Figures 2 and 3). This included the absence of a transcriptional activity of the candidate genes MA_53072g0010, MA_84091g0010, MA_19645g0010 and MA_25569g0010. The repression of the putative bHLH transcription factor MA_14352g0010 in response to *H. parviporum* was not detected in the qPCR experiment. The qPCR did confirm that *PaLAC5* (MA_97119g0010 and MA_97119g0020) is strongly and specifically upregulated in close proximity to the *H. parviporum* inoculation site (Figure 3d). Two of the candidate genes linked to the SWG QTLs with detected expression in the Norwood database, MA_17884g0010 and MA_53835g0010, were shown to be induced in response to *H. parviporum* compared to the control (Figure 3f,g). None of the tested candidate genes, including MA_17884g0010 and MA_53835g0010, were differentially expressed between H. parviporum inoculation and wounding in sapwood in early interactions (Table S4 and Method Section in Data S1).

5 | DISCUSSION

5.1 | Twelve distinct QTLs for resistance to *H. parviporum* identified by GWAS

In this study, the GWAS identified 11 significant associations across the two traits for *H. parviporum* resistance. QTLs for LL and SWG traits detected in the GWAS explained similar fractions of the observed phenotypic variation, as in the IM-based QTL study by Lind et al. (2014). However, the narrow-sense heritability of the phenotypic traits was considerably higher among the 466 Norway spruce half-sib families than in the single family used in the IM-based QTL study, 0.42 compared to 0.11 for SWG (Arnerup et al., 2010; Chen et al., 2018; Lind et al., 2014). The fact that the Norway spruce genome v 1.0 assembly was highly fragmented comprising >10 million scaffolds over 500 bp (Bernhardsson et al., 2019; Nystedt et al., 2013) made it difficult to evaluate how the QTLs identified by GWAS relate to the previously identified QTLs (Lind et al., 2014), or to each other. However, the newly published ultra-dense genetic map (Bernhardsson et al., 2019) showed that five of the QTLs were independent from the other QTL regions as they were found in different linkage groups. Only one of the QTL regions that was identified in the linkage map

FIGURE 3 Expression profile of candidate genes for *H. parviporum* resistance in response to *H. parviporum* inoculation (Hp) and wounding (W) at seven dpi proximally (0–5 mm from the inoculation site, indicated by the letter "A" in, e.g., the treatment "Hp_A") and distally (10–15 mm from the inoculation site, indicated by the letter "C") in the Norway spruce clone S21K7820222 as detected by qPCR. Candidates a–d are associated with the trait LL and candidate genes e–g with trait the SWG. The floating bars in the graphs indicate min and max values, the line indicates mean, and different letters over the bars in the graph indicate significant differences in the statistical analyses (N = 4)
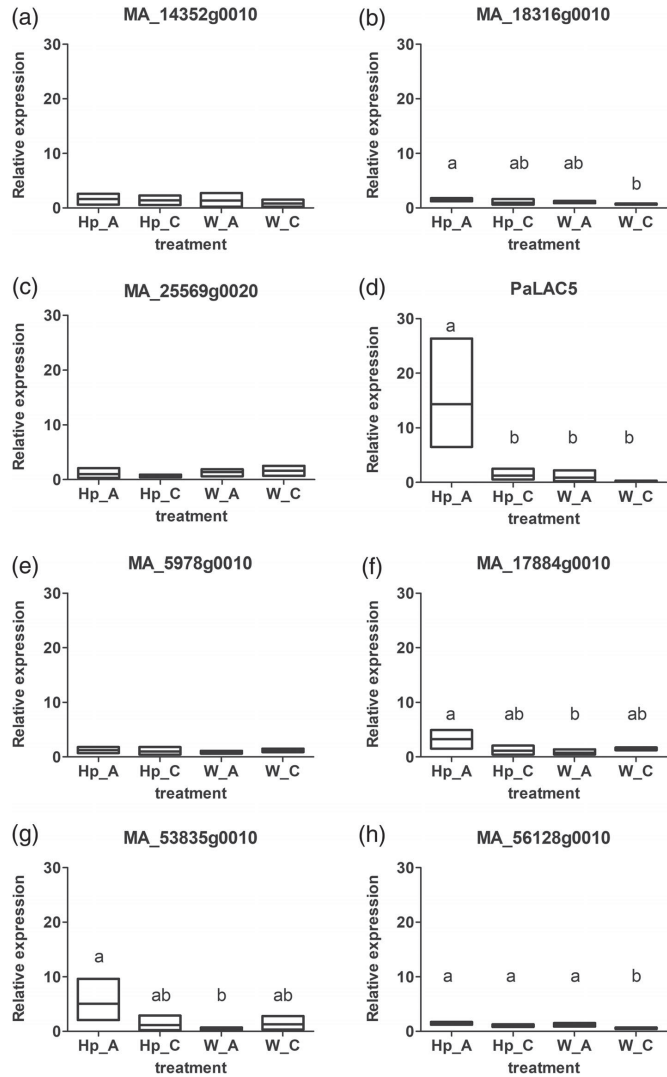
may possibly coincide with a previously identified resistance QTLs (Lind et al., 2014). The SNP MA_53835_9763 is positioned within 4 cM from a probe in the confidence region for the trait infection prevention (IP) on LG 11 (Lind et al., 2014; Chaudhary et al., submitted manuscript). Thus, the possibility that these markers target the same genomic region cannot be excluded, although it is not very likely given the short LD. Overall, the GWAS returned 11 new potential markers for resistance to *H. parviporum* in Norway spruce that could be used to aid selection in breeding programmes.

5.2 | Candidate genes have orthologues whose genetic variation is associated with the control of the responses to multiple stresses

Three of the candidate genes identified in the GWAS, MA_17884g0010, MA_5978g0020 and MA_18316g0010, have *Arabidopsis* orthologues AT3G20000.1, AT5G17640.1 and AT1G34220.2, respectively. These orthologues hold QTLs for responses to multiple stresses (Kawa et al., 2016; Thoen et al., 2017).

The candidate gene MA_18316g0010 was associated with control of lesion length in the inner bark and it was upregulated in response to *H. parviporum* inoculation compared to wounding alone, both proximally and distally. The *Arabidopsis* orthologue AT1G34220.2 encodes IST1-LIKE1 (ISTL1), a protein predicted to be the *Arabidopsis* homologue of yeast IST (Buono et al., 2016). ISTL1 is a regulator of the multivesicular bodies (MVB) pathway in which ubiquitinated and endocytically internalized membrane proteins are degraded (C. Gao, Zhuang, Shen, & Jiang, 2017). ISTL1, in interaction with LIP5 (LYST INTERACTING PROTEIN 5, AT4G26750), is essential for normal plant growth and repression of spontaneous cell death (Buono et al., 2016). The fungus *H. parviporum* is a necrotrophic pathogen and upon infection or inoculation in trees, it will create necrotic lesions in the phloem to gain access to the sapwood (Johansson & Stenlid, 1985; Lindberg & Johansson, 1991). It is, therefore, tempting to propose that the MA_18316g0010 protein fulfils the same role in the control of the cell death process as the ISTL1/LIP5 complex, MA_18316g0010 was upregulated in response to *H. parviporum* inoculation to repress cell death, a mechanism that must be integral to the LL trait. It would be interesting to test if the variation at MA_18316_3156 leads to differential accumulation of the transcript in response to *H. parviporum*.

The *Arabidopsis* orthologue to MA_17884g0010, AT3G20000.1, encodes translocase of the outer mitochondrial membrane 40, TOM40. *AtTOM40* is in LD with a QTL (Ch3:6968031) identified in a multi-trait QTL mixed models GWAS using the responses to a set of 30 biotic and abiotic stresses in 196 accessions of *Arabidopsis* (Thoen et al., 2017). TOM40 protein is the central channel forming units of the TOM complex (Hill et al., 1998). The TOM complex and the mitochondrial outer membrane play a central role in the interaction between the mitochondrion and the cytosol. It mediates the import of preproteins, the passage of small molecules and the transduction of signals between cellular compartments (Duncan, van der Merwe, Daley, & Whelan, 2013). Consequently, it is perhaps not unexpected that genetic variation associated with MA_17884g0010 and *TOM40* may influence plants responses to stress, or that MA_17884g0010 shows a ubiquitous expression in the surveyed Norway spruce tissues, with a slight upregulation in metabolically very active tissues (eg the cambium) and in response to *H. parviporum* inoculation.

5.3 | Candidate genes linked to SWG QTLs are more commonly expressed in wood

Despite the economic and ecological importance of conifers, we know surprisingly little about the genetic basis of resistance to decay pathogens compared to canker-forming pathogens in conifers (Kinloch, Sniezko, & Dupper, 2003; Liu et al., 2017; Sniezko, Smith, Liu, & Hamelin, 2014). Examining the regions under selection in response to given pathogens or stressors, identifying and testing candidate genes, can lead to better understanding of the interaction between the host and the pathogen (Liu et al., 2017; Martin, Rönnberg-Wästljung, Stenlid, & Samils, 2016; Nemesio-Gorriz et al., 2016; Thoen et al., 2017).

Under the expectation that candidate genes linked to the control of SWG are involved in processes shaping the cell wall or in production of, for example, specialized metabolites in wood (Oliva et al., 2015; Popoff, Theander, & Johansson, 1975; Stenlid & Johansson, 1987), we predicted that the expression of the candidate genes linked to SWG QTLs should be more commonly detected in the wood-forming tissues than the genes linked to the LL QTLs. A trend for this was observed in the NorWood database (Jokipii-Lukkari et al., 2017), although a larger number of QTLs and candidate genes for both traits studied would probably have been needed to gain conclusive evidence. It is, however, important to point out that none of the QTLs identified for SWG, or LL, coincide with the 52 QTLs for important wood quality traits in Norway spruce reported by Baison et al. (2019). An observation that is fully in agreement with the absence of significant correlations between wood quality, or growth, traits and resistance to *H. parviporum* in this material (Chen et al., 2018), suggesting that the detected SWG QTLs may be associated to distinct defence-related processes. Several of the expressed candidate genes showed their highest transcriptional activity in the cambium and expanding early wood libraries. The candidate gene MA_25569g0020, associated with LL, showed increased transcriptional activity during visual appearance of dead early wood cells in the sapwood. The transcript is also specifically expressed in the phloem in the autumn/winter (Jokipii-Lukkari et al., 2018), but it was not induced by *H. parviporum* inoculation. This points to that the role of MA_25569g0020 in resistance may be associated to the constitutive defence.

5.4 | The Norway spruce laccase *PaLAC5* responds specifically to *H. parviporum* inoculation

Two candidate genes associated with the LL trait in bark, MA_53835g0010 and *PaLAC5*, are likely to be members of the induced defence to *H. parviporum*. The Norway spruce *laccase* gene *PaLAC5* (MA_97119g0010 and MA_97119g0020) was originally isolated from lignin-producing Norway spruce suspension cultures (Koutaniemi, Malmberg, Simola, Teeri, & Kärkönen, 2015), and transcriptome analyses of these lignin-producing Norway spruce suspension cultures under different conditions suggest that *Pa*LAC5 is associated with the activation of stress associated lignin production (Laitinen et al., 2017). *PaLAC5* has a very specific spatial expression pattern in response to *H. parviporum* inoculation. It is strongly, and specifically, upregulated proximally to the *H. parviporum* inoculation site but not regulated 10 mm away from the developing necrotic lesion or in response to the wounding control. In contrast to the induction of *PaLAC5* in stress associated lignin production conditions in vitro, the transcriptional activity of *PaLAC5* is very low in sapwood (Blokhina et al., 2019; Jokipii-Lukkari et al., 2017; Laitinen et al., 2017). Therefore, *Pa*LAC5 is not likely to be associated with lignifying tracheids or ray parenchyma cells indicating that the induction of *PaLAC5* expression under lignin-forming conditions in the cell cultures is stress-associated and not directly connected to lignification processes in wood (Blokhina et al., 2019; Jokipii-Lukkari et al., 2017;

Laitinen et al., 2017). However, if *PaLAC5* would be responding to stress in general, it would likely have had an expression pattern similar to many other studied defense genes, which often show upregulation in proximal to both mechanical wounding sites and to inoculation points (Arnerup et al., 2011; Danielsson et al., 2011; Ralph et al., 2006). Instead, it showed a distinct expression pattern. Thus, it is probable that *PaLAC5* expression is associated with specific cell types or processes such as the formation of the LSZ in the bark adjacent to the inoculation site. The LSZ is characterized by deposition of phenolics and suberin, and an early development of a discernible LSZ is crucial in stopping fungal invasions (Bodles et al., 2007; Lindberg & Johansson, 1991; Solla, Tomlinson, & Woodward, 2002; Woodward et al., 2007). Recently, it was suggested that specific isoforms of peroxidase and laccases may be involved in cross-linking aromatics to form lignin-like polyphenolics in the suberin in bark (Rains, Molina, & Gardiyehewa de Silva, 2017). The expression pattern of *PaLAC5* responding to *H. parviporum* and lignin-forming conditions (Laitinen et al., 2017) clearly makes it an interesting candidate for such a role. It remains to be seen if *PaLAC5*, indeed, is involved in the LSZ formation and if genetic variation associated with *PaLAC5* influences the formation of the LSZ.

6 | CONCLUSIONS

Our large sample sizes and a relatively high number of markers allowed us to link traits to SNPs with GWAS and to identify candidate genes associated with the QTLs. These candidate genes present new insights into the interaction between Norway spruce and *H. parviporum*, such as a putative involvement of the secretory and endosomal trafficking pathways and the *laccase PaLAC5*, in the control of lesion extension in the inner bark or the potential role of mitochondrial protein import and biogenesis in controlling *H. parviporum* spread in the sapwood.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceived the study: H.W., J.S., B.K., M.R.G.G. and M.E.

Planned the study: B.K., M.R.G.G., M.E., I.V.

Performed experiments: K.L., I.V., M.E., H.C., M.S.Å., R.C., Å.O.

Analysed data: L.Z., J.B., Z.-Q.C., M.E., K.L., R.C.

Drafted the MS: M.E. and J.B.

Commented on MS: J.B., Z.-Q.C., K.L., B.K., M.E., H.C., M.S.Å., R.C., J.S., Å.O. and M.R.G.G.

Wrote the final MS: M.E., all authors read and approved the final version.

ORCID

Malin Elfstrand https://orcid.org/0000-0002-0214-5284
María Rosario García-Gil https://orcid.org/0000-0002-6834-6708

REFERENCES

Arnerup, J., Lind, M., Olson, A., Stenlid, J., & Elfstrand, M. (2011). The pathogenic white-rot fungus *Heterobasidion parviporum* triggers non-specific defence responses in the bark of Norway spruce. *Tree Physiology*, 31(11), 1262–1272. https://doi.org/10.1093/treephys/tpr113

Arnerup, J., Swedjemark, G., Elfstrand, M., Karlsson, B., & Stenlid, J. (2010). Variation in growth of *Heterobasidion parviporum* in a full-sib family of *Picea abies*. *Scandinavian Journal of Forest Research*, 25(2), 106–110. https://doi.org/10.1080/02827581003730799

Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43(1), 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Baison, J., Vidalis, A., Zhou, L., Chen, Z.-Q., Li, Z., Sillanpää, M. J., ... García-Gil, M. R. (2019). Genome-wide association study (GWAS) identified novel candidate loci affecting wood formation in Norway spruce. *The Plant Journal*, 100(1), 83–100. https://doi.org/10.1111/tpj.14429

Bartholomé, J., Bink, M. C. A. M., van Heerwaarden, J., Chancerel, E., Boury, C., Lesur, I., ... Plomion, C. (2016). Linkage and association mapping for two major traits used in the maritime pine breeding program: Height growth and stem straightness. *PLoS One*, 11(11), e0165323. https://doi.org/10.1371/journal.pone.0165323

Bernhardsson, C., Vidalis, A., Wang, X., Scofield, D. G., Schiffthaler, B., Baison, J., ... Ingvarsson, P. K. (2019). An ultra-dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway spruce (*Picea abies*). *G3-Genes Genomes Genetics*, 9(5), 1623–1632. https://doi.org/10.1534/g3.118.200840

Blokhina, O., Laitinen, T., Hatakeyama, Y., Delhomme, N., Paasela, T., Zhao, L., ... Fagerstedt, K. (2019). Ray parenchymal cells contribute to lignification of tracheids in developing xylem of Norway spruce. *Plant Physiology*, 181(4), 1552–1572. https://doi.org/10.1104/pp.19.00743

Bodles, W. J. A., Beckett, E., & Woodward, S. (2007). Responses of Sitka spruce of different genetic origins to inoculation with *Heterobasidion annosum*: Lesion lengths, fungal growth and development of the lignosuberized boundary zone. *Forest Pathology*, 37(3), 174–186. https://doi.org/10.1111/j.1439-0329.2007.00494.x

Bühlmann, P., Kalisch, M., & Meier, L. (2014). High-dimensional statistics with a view toward applications in biology. *Annual Review of Statistics and its Application*, 1(1), 255–278. https://doi.org/10.1146/annurev-statistics-022513-115545

Buono, R. A., Paez-Valencia, J., Miller, N. D., Goodman, K., Spitzer, C., Spalding, E. P., & Otegui, M. S. (2016). Role of SKD1 regulators LIP5 and IST1-LIKE1 in endosomal sorting and plant development. *Plant Physiology*, 171(1), 251–264. https://doi.org/10.1104/pp.16.00240

Chang, S., Puryear, J., & Cairney, J. (1993). A simple and efficient method for extracting RNA from pine trees. *Plant Molecular Biology Reporter*, 11(2), 113–116.

Chen, Z.-Q., Lundén, K., Karlsson, B., Vos, I., Olson, Å., Lundqvist, S.-O., … Elfstrand, M. (2018). Early selection for resistance to *Heterobasidion parviporum* in Norway spruce is not likely to adversely affect growth and wood quality traits in late-age performance. *European Journal of Forest Research*, 137, 517–525. https://doi.org/10.1007/s10342-018-1120-5

Cubbage, F. W., Pye, J. M., Holmes, T. P., & Wagner, J. E. (2000). An economic evaluation of fusiform rust protection research. *Southern Journal of Applied Forestry*, 24(2), 77–85. https://doi.org/10.1093/sjaf/24.2.77

Danielsson, M., Lunden, K., Elfstrand, M., Hu, J., Zhao, T., Arnerup, J., … Stenlid, J. (2011). Chemical and transcriptional responses of Norway spruce genotypes with different susceptibility to *Heterobasidion spp.* infection. *BMC Plant Biology*, 11, 154.

Duncan, O., van der Merwe, M. J., Daley, D. O., & Whelan, J. (2013). The outer mitochondrial membrane in higher plants. *Trends in Plant Science*, 18(4), 207–217. https://doi.org/10.1016/j.tplants.2012.12.004

Gao, C., Zhuang, X., Shen, J., & Jiang, L. (2017). Plant ESCRT complexes: Moving beyond endosomal sorting. *Trends in Plant Science*, 22(11), 986–998. https://doi.org/10.1016/j.tplants.2017.08.003

Gao, H., Wu, Y., Li, J., Li, H., Li, J., & Yang, R. (2014). Forward LASSO analysis for high-order interactions in genome-wide association study. *Briefings in Bioinformatics*, 15(4), 552–561. https://doi.org/10.1093/bib/bbt037

Garbelotto, M., & Gonthier, P. (2013). Biology, epidemiology, and control of *Heterobasidion* species worldwide. *Annual Review of Phytopathology*, 51(1), 39–59. https://doi.org/10.1146/annurev-phyto-082712-102225

Hall, D., Hallingbäck, H. R., & Wu, H. X. (2016). Estimation of number and size of QTL effects in forest tree traits. *Tree Genetics & Genomes*, 12(6), 110. https://doi.org/10.1007/s11295-016-1073-0

Hill, K., Model, K., Ryan, M. T., Dietmeier, K., Martin, F., Wagner, R., & Pfanner, N. (1998). Tom40 forms the hydrophilic channel of the mitochondrial import pore for preproteins. *Nature*, 395(6701), 516–521. https://doi.org/10.1038/26780

Johansson, M., & Stenlid, J. (1985). Infection of roots of Norway spruce (*Picea abies*) by *Heterobasidion annosum*. 1. Initial reactions in sapwood by wounding and infection. *European Journal of Forest Pathology*, 15(1), 32–45.

Johansson, M., & Theander, O. (1974). Changes in sapwood of roots of Norway spruce , attecked by *Fomes annosus*. Part 1. *Physiologia Plantarum*, 30(3), 218–225.

Jokipii-Lukkari, S., Delhomme, N., Schiffthaler, B., Mannapperuma, C., Prestele, J., Nilsson, O., … Tuominen, H. (2018). Transcriptional roadmap to seasonal variation in wood formation of Norway spruce. *Plant Physiology*, 176(4), 2851–2870. https://doi.org/10.1104/pp.17.01590

Jokipii-Lukkari, S., Sundell, D., Nilsson, O., Hvidsten, T. R., Street, N. R., & Tuominen, H. (2017). NorWood: A gene expression resource for evo-devo studies of conifer wood development. *New Phytologist*, 216(2), 482–494. https://doi.org/10.1111/nph.14458

Karlsson, B., & Swedjemark, G. (2006). Genotypic variation in natural infection frequency of *Heterobasidion spp.* in a *Picea abies* clone trial in southern Sweden. *Scandinavian Journal of Forest Research*, 21(2), 108–114. https://doi.org/10.1080/02827580500529969

Kawa, D., Julkowska, M. M., Sommerfeld, H. M., ter Horst, A., Haring, M. A., & Testerink, C. (2016). Phosphate-dependent root system architecture responses to salt stress. *Plant Physiology*, 172(2), 690. https://doi.org/10.1104/pp.16.00712

Kinloch, B. B., Sniezko, R. A., & Dupper, G. E. (2003). Origin and distribution of Cr2, a gene for resistance to white pine blister rust in natural populations of western white pine. *Phytopathology*, 93(6), 691–694. https://doi.org/10.1094/phyto.2003.93.6.691

Koutaniemi, S., Malmberg, H. A., Simola, L. K., Teeri, T. H., & Kärkönen, A. (2015). Norway spruce (*Picea abies*) laccases: Characterization of a laccase in a lignin-forming tissue culture. *Journal of Integrative Plant Biology*, 57(4), 341–348. https://doi.org/10.1111/jipb.12333

Laitinen, T., Morreel, K., Delhomme, N., Gauthier, A., Schiffthaler, B., Nickolov, K., … Kärkönen, A. (2017). A key role for apoplastic H_2O_2 in Norway spruce phenolic metabolism. *Plant Physiology*, 174(3), 1449–1475. https://doi.org/10.1104/pp.17.00085

Li, Z., Hallingbäck, H. R., Abrahamsson, S., Fries, A., Andersson Gull, B., Sillanpää, M. J., & García-Gil, M. R. (2014). Functional multi-locus QTL mapping of temporal trends in scots pine wood traits. *G3-Genes Genomes Genetics*, 4(12), 2365–2379. https://doi.org/10.1534/g3.114.014068

Li, Z., & Sillanpää, M. J. (2015). Dynamic quantitative trait locus analysis of plant phenomic data. *Trends in Plant Science*, 20(12), 822–833. https://doi.org/10.1016/j.tplants.2015.08.012

Lind, M., Källman, T., Chen, J., Ma, X.-F., Bousquet, J., Morgante, M., … Stenlid, J. (2014). A *Picea abies* linkage map based on SNP markers identifies QTLs for four aspects of resistance to *Heterobasidion parviporum* infection. *PLoS ONE*, 9(7), e101049. https://doi.org/10.1371/journal.pone.0101049

Lindberg, M., & Johansson, M. (1991). Growth of *Heterobasidion annosum* through bark of *Picea abies*. *European Journal of Forest Pathology*, 21(6-7), 377–388. https://doi.org/10.1111/j.1439-0329.1991.tb00775.x

Liu, J.-J., Sniezko, R. A., Zamany, A., Williams, H., Wang, N., Kegley, A., … Sturrock, R. N. (2017). Saturated genic SNP mapping identified functional candidates and selection tools for the *Pinus monticola* Cr2 locus controlling resistance to white pine blister rust. *Plant Biotechnology Journal*, 15(9), 1149–1162. https://doi.org/10.1111/pbi.12705

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, 25(4), 402–408.

Martin, T., Rönnberg-Wästljung, A.-C., Stenlid, J., & Samils, B. (2016). Identification of a differentially expressed TIR-NBS-LRR gene in a major QTL associated to leaf rust resistance in salix. *PLoS One*, 11(12), e0168776. https://doi.org/10.1371/journal.pone.0168776

Mullick, D. B. (1977). The non-specific nature of defense in bark and wood during wounding, insect and pathogen attack. In F. A. Loewus & V. C. Runeckles (Eds.), *The structure, biosynthesis, and degradation of wood* (pp. 395–441). Boston, MA: Springer.

Namroud, M.-C., Guillet-Claude, C., Mackay, J., Isabel, N., & Bousquet, J. (2010). Molecular evolution of regulatory genes in spruces from different species and continents: Heterogeneous patterns of linkage disequilibrium and selection but correlated recent demographic changes. *Journal of Molecular Evolution*, 70(4), 371–386. https://doi.org/10.1007/s00239-010-9335-1

Neale, D. B., & Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends in Plant Science*, 9(7), 325–330. https://doi.org/10.1016/j.tplants.2004.05.006

Nemesio-Gorriz, M., Hammerbacher, A., Ihrmark, K., Kallman, T., Olson, A., Lascoux, M., … Elfstrand, M. (2016). Different alleles of a gene encoding leucoanthocyanidin reductase (PaLAR3) influence resistance against the fungus *Heterobasidion parviporum* in *Picea abies*. *Plant Physiology*, 171(4), 2671–2681. https://doi.org/10.1104/pp.16.00685

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., … Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497(7451), 579–584. https://doi.org/10.1038/nature12211

Oliva, J., Bendz-Hellgren, M., & Stenlid, J. (2011). Spread of *Heterobasidion annosum* s.s. and *Heterobasidion parviporum* in *Picea abies* 15 years after stump inoculation. *FEMS Microbiology Ecology*, 75(3), 414–429. https://doi.org/10.1111/j.1574-6941.2010.01020.x

Oliva, J., Rommel, S., Fossdal, C. G., Hietala, A. M., Nemesio-Gorriz, M., Solheim, H., & Elfstrand, M. (2015). Transcriptional responses of Norway spruce (*Picea abies*) inner sapwood against Heterobasidion parviporum. *Tree Physiology*, 35(9), 1007–1015. https://doi.org/10.1093/treephys/tpv063

Pautasso, M., Schlegel, M., & Holdenrieder, O. (2015). Forest health in a changing world. *Microbial Ecology*, 69(4), 826–842. https://doi.org/10.1007/s00248-014-0545-8

Petit, R. J., & Hampe, A. (2006). Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics*, 37(1), 187–214. https://doi.org/10.1146/annurev.ecolsys.37.091305.110215

Popoff, T., Theander, O., & Johansson, M. (1975). Changes in sapwood of roots of Norway spruce, attacked by *Fomes annosus*. *Physiologia Plantarum*, 34(4), 347–356. https://doi.org/10.1111/j.1399-3054.1975.tb03851.x

Porth, I., Hamberger, B., White, R., & Ritland, K. (2011). Defense mechanisms against herbivory in *Picea*: Sequence evolution and expression regulation of gene family members in the phenylpropanoid pathway. *BMC Genomics*, 12(608), 1–26. https://doi.org/10.1186/1471-2164-12-608

R Studio Team. (2015). *RStudio: Integrated development for R*. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com

Rains, M. K., Molina, I., & Gardiyehewa de Silva, N. D. (2017). Reconstructing the suberin pathway in poplar by chemical and transcriptomic analysis of bark tissues. *Tree Physiology*, 38(3), 340–361. https://doi.org/10.1093/treephys/tpx060

Ralph, S. G., Yueh, H., Friedmann, M., Aeschliman, D., Zeznik, J. A., Nelson, C. C., ... Bohlmann, J. (2006). Conifer defence against insects: Microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale changes of the host transcriptome. *Plant, Cell & Environment*, 29(8), 1545–1570.

Redfern, D. B., & Stenlid, J. (1998). Spore dispersal and infection. In S. Woodward, J. Stenlid, R. Karjalainen, & A. Hüttermann (Eds.), *Heterobasidion annosum: Biology, ecology, impact and control* (pp. 105–124). London, England: CAB International.

Shain, L. (1971). Response of sapwood of Norway spruce to infectopn by *Fomes annosus*. *Phytopathology*, 61(3), 301–307.

Skrøppa, T., Solheim, H., & Steffenrem, A. (2015). *Genetic variation, inheritance patterns and parent–offspring relationships after artificial inoculations with Heterobasidion parviporum and Ceratocystis polonica in Norway spruce seed orchards and progeny tests*, 49.

Sniezko, R. A., Smith, J., Liu, J.-J., & Hamelin, R. C. (2014). Genetic resistance to fusiform rust in southern pines and white pine blister rust in white pines—A contrasting tale of two rust pathosystems—Current status and future prospects. *Forests*, 5(9), 2050–2083.

Solla, A., Tomlinson, F., & Woodward, S. (2002). Penetration of *Picea sitchensis* root bark by *Armillaria mellea*, *Armillaria ostoyae* and *Heterobasidion annosum*. *Forest Pathology*, 32(1), 55–70. https://doi.org/10.1046/j.1439-0329.2002.00265.x

Steffenrem, A., Solheim, H., & Skrøppa, T. (2016). Genetic parameters for wood quality traits and resistance to the pathogens *Heterobasidion parviporum* and *Endoconidiophora polonica* in a Norway spruce breeding population. *European Journal of Forest Research*, 135(5), 815–825. https://doi.org/10.1007/s10342-016-0975-6

Stenlid, J., & Johansson, M. (1987). Infection of roots of Norway spruce (*Picea abies*) by *Heterobasidion annosum*. *European Journal of Forest Pathology*, 17(4-5), 217–226. https://doi.org/10.1111/j.1439-0329.1987.tb01019.x

Stenlid, J., & Swedjemark, G. (1988). Differential growth of S- and P-isolates of *Heterobasidion annosum* in *Picea abies* and *Pinus sylvestris*. *Transactions of the British Mycological Society*, 90(2), 209–213.

Thoen, M. P. M., Davila Olivas, N. H., Kloth, K. J., Coolen, S., Huang, P.-P., Aarts, M. G. M., ... Dicke, M. (2017). Genetic architecture of plant stress resistance: Multi-trait genome-wide association mapping. *New Phytologist*, 213(3), 1346–1362. https://doi.org/10.1111/nph.14220

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, 31(1), 46–53. https://doi.org/10.1038/nbt.2450

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7(3), 562–578. https://doi.org/10.1038/nprot.2012.016

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... Pachter, L. (2014). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks (vol 7, pg 562, 2012). *Nature Protocols*, 9(10), 2513–2513. https://doi.org/10.1038/nprot1014-2513a

Vestman, D., Larsson, E., Uddenberg, D., Cairney, J., Clapham, D., Sundberg, E., & von Arnold, S. (2011). Important processes during differentiation and early development of somatic embryos of Norway spruce as revealed by changes in global gene expression. *Tree Genetics & Genomes*, 7(2), 347–362. https://doi.org/10.1007/s11295-010-0336-4

Vidalis, A., Scofield, D. G., Neves, L. G., Bernhardsson, C., García-Gil, M. R., & Ingvarsson, P. (2018). Design and evaluation of a large sequence-capture probe set and associated SNPs for diploid and haploid samples of Norway spruce (*Picea abies*). *bioRxiv*. Retrieved from https://doi.org/10.1101/291716

Woodward, S., Bianchi, S., Bodles, W. J. A., Beckett, L., & Michelozzi, M. (2007). Physical and chemical responses of Sitka spruce (*Picea sitchensis*) clones to colonization by *Heterobasidion annosum* as potential markers for relative host susceptibility. *Tree Physiology*, 27(12), 1701–1710. https://doi.org/10.1093/treephys/27.12.1701

Woodward, S., & Pocock, S. (1996). Formation of the ligno-suberized barrier zone and wound periderm in four species of European broad-leaved trees. *European Journal of Forest Pathology*, 26(2), 97–105. https://doi.org/10.1111/j.1439-0329.1996.tb00714.x

Woodward, S., Stenlid, J., Karjalainen, R., & Hüttermann, A. (1998). Preface. In S. Woodward, J. Stenlid, R. Karjalainen, & A. Hüttermann (Eds.), *Heterobasidion annosum: Biology, ecology, impact and control* (pp. xi–xii). London, England: CAB International.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418–1429. https://doi.org/10.1198/016214506000000735

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

IV

ORIGINAL ARTICLE

Genotypic variation in Norway spruce correlates to fungal communities in vegetative buds

Malin Elfstrand[1] | Linghua Zhou[2] | John Baison[2] | Åke Olson[1] | Karl Lundén[1] | Bo Karlsson[3] | Harry X. Wu[2] | Jan Stenlid[1] | M. Rosario García-Gil[2]

[1]Uppsala Biocentre, Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden

[2]Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, Umeå, Sweden

[3]Skogforsk, Svalöv, Sweden

Correspondence
Malin Elfstrand, Uppsala Biocentre, Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden.
Email: Malin.Elfstrand@slu.se

Funding information
Svenska Forskningsrådet Formas, Grant/Award Number: 2015-00081; Stiftelsen för Strategisk Forskning, Grant/Award Number: RBP14-0040

Abstract

The taxonomically diverse phyllosphere fungi inhabit leaves of plants. Thus, apart from the fungi's dispersal capacities and environmental factors, the assembly of the phyllosphere community associated with a given host plant depends on factors encoded by the host's genome. The host genetic factors and their influence on the assembly of phyllosphere communities under natural conditions are poorly understood, especially in trees. Recent work indicates that Norway spruce (*Picea abies*) vegetative buds harbour active fungal communities, but these are hitherto largely uncharacterized. This study combines internal transcribed spacer sequencing of the fungal communities associated with dormant vegetative buds with a genome-wide association study (GWAS) in 478 unrelated Norway spruce trees. The aim was to detect host loci associated with variation in the fungal communities across the population, and to identify loci correlating with the presence of specific, latent, pathogens. The fungal communities were dominated by known Norway spruce phyllosphere endophytes and pathogens. We identified six quantitative trait loci (QTLs) associated with the relative abundance of the dominating taxa (i.e., top 1% most abundant taxa). Three additional QTLs associated with colonization by the spruce needle cast pathogen *Lirula macrospora* or the cherry spruce rust (*Thekopsora areolata*) in asymptomatic tissues were detected. The identification of the nine QTLs shows that the genetic variation in Norway spruce influences the fungal community in dormant buds and that mechanisms underlying the assembly of the communities and the colonization of latent pathogens in trees may be uncovered by combining molecular identification of fungi with GWAS.

KEYWORDS
cherry spruce rust, peroxidase, phenology, Picea abies, Rhizosphaera kalkhoffii, Sydowia polyspora

---

Linghua Zhou and John Baison contributed equally.

1 | INTRODUCTION

At any given time during their life cycle, trees are colonized by a wide range of microbes, including fungi. These fungi may reside both epiphytically, on the surface of the tissue, and endophytically, within the host tissue without causing any visible symptoms (Arnold, Henk, Eells, Lutzoni, & Vilgalys, 2007; Rajala et al., 2013; Rodriguez, White, Arnold, & Redman, 2009; Sieber, 2007). The epi- or endophytic fungi associated with a tree have varying ecological roles, such as being mutualistic symbionts (latent) pathogens or facultative saprotrophs (Rajala, Velmala, Vesala, Smolander, & Pennanen, 2014; Saikkonen, 2007). Pathogenic fungi cause damage to the host that can reduce host growth and fitness, such as by decreasing photosynthetic capacity, and causing premature leaf shed or lesion formation (Hanso & Drenkhan, 2012; Pan et al., 2018). Highly diverse fungal communities have been reported in studies on current-year Norway spruce needles, that is those that flushed during the sampling season (Menkis, Marciulynas, Gedminas, Lynikiene, & Povilaitiene, 2015; Nguyen, Boberg, Ihrmark, Stenström, & Stenlid, 2016). It has been shown that the needle endophyte community in Norway spruce (*Picea abies* L., Karst) varies among genotypes (Rajala et al., 2013, 2014). This would indicate that apart from the presence of fungi with the capacity to colonize the host, environmental and climatic factors (Eusemann et al., 2016; Menkis et al., 2015; Millberg, Boberg, & Stenlid, 2015; Moler & Aho, 2018; Nguyen et al., 2016; Rodriguez et al., 2009), the assembly of the phyllosphere community associated with a given tree will also depend on factors encoded by the host's genotype (Cordier, Robin, Capdevielle, Desprez-Loustau, & Vacher, 2012; Rajala et al., 2013, 2014). Although it is hypothesized that the plant-inhabiting community is synergistically determined by environmental and host genetic factors (Horton et al., 2014; Rajala et al., 2013; Terhonen, Blumenstein, Kovalchuk, & Asiegbu, 2019), the role of specific host genetic variation under natural conditions is poorly understood. Reports on specific host genetic variation that affects the phyllosphere community composition come from annual plant species and deal primarily with bacterial colonization (Horton et al., 2014; Roman-Reyna et al., 2019; Wallace, Kremling, Kovar, & Buckler, 2018). In a study of 196 accessions of *Arabidopsis thaliana*, Horton et al. (2014) showed that both the presence/absence and abundance of fungal species in the associated phyllosphere communities are influenced by the *Arabidopsis* genotype, but for only the more abundant species in the study. Genome wide association study (GWAS) results have implicated a role of both defence responses and cell-wall integrity in the assembly of *Arabidopsis* phyllosphere fungal communities (Horton et al., 2014). However, hitherto there have been no reports on specific genetic variants affecting phyllosphere fungal community composition in perennial plant species, such as trees. These have a more complex architecture (relative to most annual plants) and are exposed to their changing biotic and abiotic environment over many consecutive growth cycles. These interactions between genetics, environment and time are likely to significantly influence the structure of the phyllosphere community of a host tree.

Conifers dominate the boreal forests in the Northern Hemisphere (Farjon & Page, 1999). Owing to their often large population sizes, outbreeding mating systems and efficient gene flow (wind pollination), conifers are characterized by high levels of heterozygosity and intraspecific diversity, which is reflected in low levels of genetic differentiation between populations (Savolainen, Pyhäjärvi, & Knürr, 2007). Norway spruce is one of the most important conifer species in Europe, both ecologically and economically. Together with Scots pine (*Pinus sylvestris* L.), it essentially makes up the continuous boreal forests of the continent. The fungal community composition of Norway spruce needles has been reported to change along a latitudinal gradient on a continental scale as well as between individual genotypes of Norway spruce within a stand (Nguyen et al., 2016; Rajala et al., 2013, 2014), a pattern which may be considered consistent with horizontal transfer of the fungi, and possibly also with the high intraspecific diversity in conifers (Prunier, Verta, & MacKay, 2016).

The development of high-throughput sequencing (HTS) methods paved the way for the generation of the first draft assembly of the Norway spruce genome (Nystedt et al., 2013). The availability of the genome sequence for Norway spruce has opened new possibilities for the development of genetic markers to produce highly resolved genotypes of an individual tree (Vidalis et al., 2018) and to conduct GWAS (Baison et al., 2019). Similarly, sequencing of the ITS (internal transcribed spacer) region with HTS methods has allowed mycologists to describe fungal communities and community dynamics in various ecosystems in greater detail than before, advancing functional understanding of various ecological processes and phylogenetic relationships (Clemmensen et al., 2013; Kubartová, Ottosson, Dahlberg, & Stenlid, 2012; Rosling et al., 2011; Seena & Monroy, 2016; Tedersoo et al., 2014; Voříšková & Baldrian, 2012). The combination of ITS sequencing of phyllosphere fungi with the recently available genotyping resources in Norway spruce in an association study may provide insights in to how tree phyllosphere communities assemble, through the identification of specific fungal taxa and with genetic variants associated with general shifts, in the tree phyllosphere communities. For instance, the communities of seemingly healthy needles often include known (e.g., *Lirula macrospora* or *Rhizosphaera kalkhoffii*) or suspected (e.g., *Phoma herbarum* and *Sydowia polyspora*) needle pathogens (Menkis et al., 2015; Nguyen et al., 2016; Rajala et al., 2013, 2014). Furthermore, in a recent meta-transcriptomics study of Norway spruce tissues, similar frequencies of fungal transcripts were found both in needle and in bud samples (Delhomme et al., 2015), suggesting that vegetative buds as well as needles harbour active fungal communities, but the study provided no insights into the composition of the bud community.

Here we report the results of a study in the perennial conifer Norway spruce that combined ITS sequencing of the fungal communities associated with dormant buds with GWAS of 478 individuals to (a) describe the fungal community associated with dormant buds in a large population of unrelated trees and thus describe the abundance of possible latent pathogen colonizations of asymptomatic tissues, (b) suggest loci in the genome of Norway spruce that correlate with variation in the fungal communities across the studied

population, and (c) identify loci in the genome of Norway spruce that correlate with the presence of specific, latent, pathogens.

2 | MATERIALS AND METHODS

2.1 | Amplification, sequencing and analysis of phyllosphere fungal community

Healthy looking vegetative buds were sampled from 518 trees in the southern Swedish Norway spruce breeding archives, located at Ekebo and Maltesholm. The trees in the breeding archive were planted approximately 35 years ago with 7 m between each tree and 5 m between rows. Buds were collected by hand from exposed branches about 2 m above ground. The collected buds were placed into labelled Ziploc plastic bags. In the field, the samples were stored in sterox boxes filled with cooling blocks. After each day of field work the samples were transferred to −20°C for long-term storage. Total genomic DNA was extracted from approximately five buds per tree using the Qiagen Plant DNA extraction kit (extraction details as described by Baison et al., 2019).

DNA samples were amplified separately using the primer pair gITS7 (5′-xxxxxxxxGTGARTCATCGARTCTTTG-3′) and ITS4 (5′-xxxxxxxxTCCTCCGCTTATTGATATGC-3′) (Ihrmark et al., 2012) containing 8-bp sample identification barcodes denoted by x, as previously described (Clemmensen, Ihrmark, Durling, & Lindahl, 2016). Each sample was amplified using unique barcode combinations. The gITS7–ITS4 amplicon amplifies the ITS2 region, which provides good species resolution capacity and can be sequenced throughout the entire length with available HTS technologies (Clemmensen et al., 2016; Ihrmark et al., 2012).

Prior to the amplification of the ITS2 region, DNA quantification was performed using the Qubit ds DNA Broad Range Assay Kit (ThermoFischer) and samples were diluted when needed ensuring sample concentration in the range of 1–10 ng/µl. PCR amplifications were done according to Clemmensen et al. (2016). In brief 50-µl reactions with a final concentration of 1× DreamTaq Green Buffer, 200 µm dNTPs, 750 µm MgCl$_2$, 1.25 µl DreamTaq polymerase (ThermoFischer Scientific), 0.3 µм Tagged gITS7/ITS4 primer mix and approximately 2.5 ng of template DNA were run with the following cycle conditions: 5 min at 94°C, 27 cycles of 30 s at 94°C, 30 s at 56°C and 30 s 72°C, and a final 7 min at 72°C; blanks were run as negative controls. The resulting amplicons were purified with the AMPure kit (Beckman Coulter). The concentration of purified PCR products was determined fluorometrically using the Qubit ds DNA Broad Range Assay Kit (ThermoFischer), and equimolar mixes of PCR products from 88 samples, including the blanks, were pooled, creating six sample pools that were used for PacBio SMRT circular consensus sequencing using eight SMRT cells per pool. Construction of the sequencing library and sequencing were carried out by NGI SciLifeLab.

The sequences generated were subjected to quality control and clustering in the scata NGS sequencing pipeline (http://scata.mykop

at.slu.se). Quality filtering of the sequences included the removal of short sequences (<200 bp), sequences with low read quality (any base in the sequence that has a PHRED score < 10) and primer dimers; homopolymers were collapsed to 3 bp before clustering. Sequences that were missing a tag or primer were excluded. The primer and sample tags were then removed from the sequence, but information on the sequence association with the sample was stored as metadata. The sequences were then clustered into different operational taxonomic units (OTUs) that essentially correspond to the species level by single-linkage clustering based on 98.5% similarity. The most common genotype (real read) for clusters was used to represent each taxon. For clusters containing two sequences, a consensus sequence was produced. The fungal taxa were taxonomically identified using the UNITE database version 7.2 (https://unite.ut.ee/index.php) and the blastn algorithm. The criteria used for identification were as follows: sequence coverage > 80%, similarity to species level 98%–100% and similarity to genus level 94%–97%. Sequences not matching these criteria were considered unidentified. To obtain further information on abundant yet unidentified sequences (i.e., suspected Norway spruce clusters), a secondary search was performed in GenBank (NCBI) using the blastn algorithm. After removal of singletons and nonfungal sequence reads, the relative abundances of each fungal cluster (OTU) in each sample were calculated. The data set was tested for PCR (polymerase chain recation) contamination by analysing the PCR blanks, and confounding factors such as variation between sequencing libraries, sites and site characteristics. From the original population samples, 473 trees growing at the Maltesholm site were selected for subsequent analysis.

The 1% most abundantly sequenced OTUs were analysed with multivariate ordination methods in Past 3.20 (Hammer, Harper, & Ryan, 2001). To identify the main drivers in the data set, a principal component analysis (PCA) was made and the first six eigenvectors from the PCA were analysed in the subsequent GWAS.

OTUs including more than 2% of the reads, corresponding to known conifer pathogens and with a presence in at least 35% of the Norway spruce samples, were selected for targeted GWAS of latent pathogens (Figure S3). Separate files with either relative abundance or presence/absence data were prepared for each of the OTUs that met the criteria for GWAS of latent pathogens and were used in trait-association mapping.

2.2 | Norway spruce genotyping and SNP annotation

Generation and evaluation of Norway spruce exome capture is described elsewhere (Vidalis et al., 2018). In brief, 478 samples from a subset of 9,000 maternal trees on which sequence capture was performed using 40,018 previously evaluated diploid probes and samples, were sequenced to an average depth of 15×. Illumina sequencing compatible libraries were amplified with 14 cycles of PCR with the probes being hybridized to a pool comprising 500 ng of eight equimolarly combined libraries following Agilent's SureSelect

Target Enrichment System (Agilent Technologies) protocol. These enriched libraries were then sequenced on an Illumina HiSeq 2500 using the 2 × 100-bp sequencing mode.

Read mapping and initial variant calling is described in detail by Baison et al. (2019). Basically, the sequence reads were aligned to the Norway spruce genome using the Burrows–Wheeler Aligner (BWA; Li & Durbin, 2010) and variant detection utilized the Genome Analysis Software Kit (GATK; DePristo et al., 2011; McKenna et al., 2010) best practice pipeline with a training subset created for application in the Variant Quality Score Recalibration (VQSR) method. Only bi-allelic single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) and "missingness" of <0.05 and >20%, respectively, were removed. For the selected set of trees a total of 178,101 SNPs passed the filtering and were used for downstream analysis. Annotation was performed using default parameters of SN-PEFF 4 (Cingolani et al., 2012) and local Norway spruce genome annotated database. Ensembl general feature format (GTF, gene sets) information was utilized to build the *Picea abies* SNPEFF database.

2.3 | Trait association mapping

Loadings on the first six axes from the PCA of the relative abundance data of the 1% largest OTUs were used for the GWAS, with each individual axis explaining at least 5% of the variance in the detected phyllosphere communities. Subsequently, the relative abundance or presence/absence data for the OTUs (OTU_5, OTU_9, OTU_15 and OTU_19) on each host genotype were used for the trait-association mapping.

The statistical LASSO model as described by Li et al. (2014) was applied to the traits associated with the detected phyllosphere communities and phyllosphere pathogens.

The LASSO model is:

$$\min_{(\alpha_0, \alpha_j)} \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \alpha_0 - \sum_{j=1}^{p} x_{ij}\alpha_j \right)^2 + \lambda \sum_{j=1}^{p} |\alpha_j|,$$

where y_i is the phenotypic value of an individual i ($i = 1, ..., n$; n is the total number of individuals), α_0 is the population mean parameter, x_{ij} is the genotypic value of individual i and marker j coded as 0, 1 and 2 for three marker genotypes AA, AB and BB, respectively, α_j is the effect of marker j ($i = 1, ..., n$; n is the total number of markers), and λ (>0) is a shrinkage tuning parameter. A fundamental idea of LASSO is to utilize the penalty function to shrink the SNP effects toward zero, and only keep a small number of important SNPs which are highly associated with the trait in the model.

The stability selection probability (SSP) of each SNP being selected by the model was applied to determine significant SNPs (Gao et al., 2014; Li & Sillanpää, 2015). For a marker to be declared significant, an SSP inclusion ratio (Frequency) was calculated for all selected SNPs for each trait and a minimum inclusion frequency of 0.56 was chosen as the most prudent cut-off. The set of markers with nonzero effects was recorded; after bootstrapping, this provides an approximation of a *p*-value. The SSP threshold used for defining significant SNPs was estimated as suggested by Meinshausen and Bühlmann (2010). The threshold is calculated based on the number of SNPs, number of individuals included in the subset and expected number of false positives. A lambda of 250 with 1,000 bootstraps and a false positive cutoff of five (5) were applied to the entire association analysis. Population structure was accounted for in all analyses by including the first five components from a PCA of the genotypic data as covariates in the LASSO model and the total variance explained by the five PCs was 19.3%. Finally, an adaptive LASSO approach (Zou, 2006) was used to determine the percentage of phenotypic variance (PVE; H^2_{QT}) of all quantitative trait loci (QTLs). The analyses were all performed with *glmnet* in RSTUDIO, R version 3.4.0 (Team, 2015), and the codes used can be found at https://github.com/RosarioGarciaLab/Norway-Spruce-Association-Mapping.

Information on putative candidate genes associated with the QTLs and the expression pattern of the candidate genes in the Norway spruce clone Z4006 were collected from the publicly available Norway spruce genome portal and *P. abies* exAtlas (https://www.congenie.org). The position of the detected QTLs in the Norway spruce genome was estimated by searching an ultradense genetic map (Bernhardsson et al., 2019) for markers derived from the same probes from which the SNP markers holding the QTLs originated.

3 | RESULTS

3.1 | Norway spruce buds are colonized by well-known phyllosphere endophytes and pathogens

After quality control, 676,375 ITS sequence reads remained. At 98.5% similarity, these sequences clustered into 4,899 OTUs, excluding 3,357 singletons. Fungal taxa were identified using the UNITE database. However, the dominant cluster (85% of reads, a spruce ITS sequence [AJ243167] and 10 other nonfungal clusters were removed from the data set leaving 4,888 OTUs [104,768 sequence reads]). After removal of the plant clusters the median read count over the samples in the data set was 148 reads. Relative abundances in each sample were calculated on the remaining OTUs. After testing and removal of confounding factors and of samples with poor amplification/sequencing success (total number of remaining fungal reads <20) the analyses were limited to the largest clusters only (Horton et al., 2014), leaving 473 Norway spruce samples and 49 OTUs (top 1%, including 80% of the fungal reads, Supporting Information S1) in the subsequent analyses. The majority of these belonged to Ascomycota, 56% belonged to Pezizomycotina alone and two OTUs (4%) represented Taphrinomycotina (Figure S2). The basidiomycete orders Agaricomycotina (25%), Pucciniomycotina (13%) and Ustilagomycotina (2%) were among the most abundant OTUs (Figure S2). Several of the most abundant OTUs represented yeasts (e.g., OTU_2 and OTU_1), and correspond to the basidiomycete yeasts *Curvibasidium cygneicollum* and *Filobasidium wieringae*,

TABLE 1 Putative taxonomic assignment for the most abundant operational taxonomic units (OTUs)

Cluster ID[a]	Cluster size[b]	Samples (%)[c]	Reference[d]	Accession[d]	Score	E-value	%
OTU_2	10,385	81.5	*Curvibasidium cygneicollum*	KY102972	573	3.00E-162	100
OTU_1	9,573	78.3	*Filobasidium wieringae*	KY103446	608	8.00E-173	100
OTU_4	8,116	85.6	Fungi	KP897174	460	2.00E-128	100
OTU_3	6,942	96.1	*Cladosporium herbarum*	MF326605	449	4.00E-125	100
OTU_5	6,855	81.7	*Sydowia polyspora*	KU837235	473	2.00E-132	100
OTU_6	5,878	81.3	*Aureobasidium pullulans*	MF687197	460	2.00E-128	100
OTU_8	4,796	76.5	*Vishniacozyma victoriae*	MF927673	433	4.00E-120	100
OTU_10	4,018	42.6	Fungi	KP897236	473	2.00E-132	100
OTU_7	3,716	71.6	*Itersonilia pannonica*	KY103613	470	3.00E-131	100
OTU_11	2,604	49.5	*Ceramothyrium* sp.	KC978733	427	2.00E-118	97.3
OTU_12	2,337	61.9	Xylariales	KP897169	494	2.00E-138	100
OTU_17	2,259	63.9	*Perusta inaequalis*	AY560007	473	2.00E-132	100
OTU_19	2,200	39.8	*Lirula macrospora*	HQ902159	392	6.00E-108	99.1
OTU_9	2,183	41.4	*Thekopsora areolata*	DQ087231	575	8.00E-163	100
OTU_14	2,178	64.1	*Tristratiperidium* sp.	KT696538	451	1.00E-125	97.0
OTU_15	2,057	57.6	*Rhizosphaera kalkhoffii*	KY003236	475	6.00E-133	100
OTU_18	1,813	77.5	*Sporobolomyces roseus*	MF102879	555	9.00E-157	100
OTU_13	1,581	71.0	Melampsoraceae	KC883505	593	2.00E-168	99.7
OTU_24	1,190	72.8	*Botrytis cinerea*	MF996364	444	2.00E-123	100
OTU_23	1,084	65.9	*Ramularia vaccinicola*	KY979777	435	1.00E-120	100
OTU_20	995	60.0	*Rhodosporidiobolus colostri*	KY104695	556	3.00E-157	100
OTU_22	908	60.0	*Exobasidium* sp.	EU692770	366	4.00E-100	96.4
OTU_25	824	60.4	Cystofilobasidiales	HM240834	571	1.00E-161	100
OTU_37	807	61.1	Leotiomycetes	KM519282	433	4.00E-120	99.2
OTU_29	785	64.5	*Taphrina carpini*	AY239215	536	3.00E-151	99.7
OTU_21	745	53.8	Fungi	KP891308	654	0	100
OTU_30	726	65.5	*Phoma herbarum*	KT355016	460	2.00E-128	100
OTU_34	690	56.0	Helotiales	AY971722	377	2.00E-103	99.5
OTU_36	647	56.6	*Dothiora* sp.	KU728514	396	5.00E-109	94.3
OTU_39	584	54.6	*Cyphellophora sessilis*	KP400571	518	1.00E-145	100
OTU_43	547	55.8	Fungi	KP897199	442	6.00E-123	100
OTU_44	539	58.2	Dothideomycetes	FR773250	433	4.00E-120	100

(Continues)

TABLE 1 (Continued)

Cluster ID[a]	Cluster size[b]	Samples (%)[c]	Reference[d]	Accession[d]	Score	E-value	%
OTU_41	459	47.7	*Phaffia* sp.	HF558647	590	3.00E-167	97.2
OTU_45	432	48.3	*Trichomerium* sp.	KP004468	444	2.00E-123	97.0
OTU_50	408	48.5	*Bannozyma* sp.	KY558343	551	3.00E-156	99.7
OTU_53	383	53.8	*Cyphellophora*	AM902002	484	1.00E-135	100
OTU_49	375	50.7	Chaetothyriales	GQ999538	518	1.00E-145	99.7
OTU_48	369	43.4	*Phallus impudicus*	UDB015413	536	3.00E-151	100
OTU_55	366	51.9	Fungi	KP897431	529	5.00E-149	100
OTU_56	362	52.3	Botryosphaeriales	KM216350	427	2.00E-118	99.2
OTU_57	355	55.4	*Dioszegia aurantiaca*	KY103346	407	2.00E-112	100
OTU_54	353	52.1	*Genolevuria* sp.	KF036585	460	2.00E-128	97.4
OTU_47	348	45.6	*Cystofilobasidium capitatum*	KY103158	593	2.00E-168	100
OTU_72	301	21.6	*Naevala* sp.	JN120378	429	5.00E-119	98.8
OTU_61	300	23.0	*Alternaria infectoria*	Y17066	468	1.00E-130	100
OTU_62	292	18.6	*Taphrinia padi*	KU134843	564	2.00E-159	99.7
OTU_60	286	18.8	Fungi	KP897203	477	2.00E-133	100
OTU_77	267	23.0	*Symmetrospora* sp.	EU002876	507	3.00E-142	97.0

[a]Cluster_ID is the identity of the assembled OTU.

[b]Cluster size is the number of reads that are associated with the OTU.

[c]Samples (%) is the frequency of the OTU presence in the 493 samples.

[d]The taxonomic assignment where "Reference" is the taxonomic assignment based on BLASTN searches in UNITE, "Accession" is the best hit, "score" and "E-value" are the BLAST score and E-values, and "%" is the percentage identity in the alignment.

respectively while OTU_6 corresponds to the common yeast-like ascomycete *Aureobasidium pullulans* (Table 1). The abundant OTU_3 is probably a member of the *Cladosporium herbarum* complex (Table 1). Other OTUs comprise well-known needle endophytes, such as *Ceramothyrium* (OTU_11) and *Perusta inaequalis* (OTU_17) (Table 1). OTU_4, one of the most abundant OTUs, is apparently a hitherto undescribed ascomycete fungus (Table 1), and is identical to the relatively abundant sequence cluster "Unidentified sp. 2168_9" reported from the Norway spruce phyllosphere (Menkis et al., 2015). Several of the abundant OTUs represent needle pathogens: OTU_9 (*Thekopsora areolata*), OTU_15 (*Rhizosphaera kalkhoffii*) and OTU_19 (*Lirula macrospora*), or latent pathogens: OTU_5 (*Sydowia polyspora*), OTU_24 (*Botrytis cinerea*) and OTU_30 (*Phoma herbarum*) (Table 1).

To reduce the number of variables for the subsequent GWAS analysis, the relative abundance data of the OTUs/species with the highest number of reads were used in a PCA to identify the most prominent drivers in the data set. The first two principal components (PCs) explained 14.9% and 12.8% of the total variance respectively in the phyllosphere community. The first and second PCs were strongly influenced by the abundance data of the undescribed ascomycete (OTU_4) and *C. cygneicollum*, while the second axis was also driven by *C. herbarum*, *S. polyspora* and OTU_10 (unclassified fungus; Figure 1a; Table 1). In addition to several of the taxa shaping the first two PCs, *F. wieringae* appeared to be an important driver on the third, and *T. areolata* on the fourth (Figure 1b; Table 1). PCs 5 and 6 explained 7.8% and 6.3% of the variance, respectively but the axes were shaped by mostly different OTUs than the first axis, such as *A. pullulans*, *T. areolata*, *R. kalkhoffii*, *Ceramothyrium* sp. and *L. macrospora* (Figure 1c; Table 1).

3.2 | Trait association-mapping

Using the loadings on the first six PCs of the PCA (Figure 1) for GWAS, six QTLs associated with the fungal community that determine the PCs were detected (Table 2). The contigs of four of these QTLs could be identified on different linkage groups in an ultradense Norway spruce genetic linkage map (Bernhardsson et al., 2019; Table S4), indicating that the QTLs are located at different positions in the genome. The three SNPs that were significantly associated with PC3 are positioned in or adjacent to four Norway spruce gene models: MA_24477g0010, MA_31029g0010, MA_10428833g0010 and MA_10428833g0020 (Table 3). The gene ontology (GO) terms GO:0007264 (small GTPase mediated signal transduction) and GO:0006886 (intracellular protein transport) are associated with the gene model MA_24477g0010 that is similar to the *Arabidopsis* ADP-ribosylation factor A1F gene (Table 3). MA_10428833g0020, encoding a putative ubiquitin-protein ligase, is associated with the terms GO:0006817 (phosphate ion transport), GO:0016036 (cellular response to phosphate starvation), GO:0010337 (regulation of salicylic acid metabolic process), GO:0080021 (response to benzoic acid) and GO:0009626 (plant-type hypersensitive response). Neither of the other two gene models have any functional information

connected to them in the Norway spruce genome portal. However, MA_31029g0010 is homologous to AT3G59340, a putative solute carrier family 35 protein which is reported to be located in *Arabidopsis* chloroplasts. Inspection of the expression patterns of MA_24477g0010, MA_31029g0010, MA_10428833g0010 and MA_10428833g0020 reported in the *Picea abies* exAtlas (Nystedt et al., 2013) suggest that the candidate genes may be highly expressed in bark phloem, cambium and primary xylem (samples Z4006TR12, Z4006TR24, Z4006TR25 and Z4006TR19) and that relatively high levels of MA_31029g0010 transcript accumulate in vegetative shoots, developing buds, pine apple galls and immature female cones (Figure 2). Interestingly, the gene model MA_208236g0010 associating with PC5 showed an expression pattern reminiscent of MA_31029g0010 with relatively high transcript expression accumulating in vegetative shoots and developing buds (Figure 2). There is no functional information for MA_208236g0010 in the Norway spruce genome portal but the putative *Arabidopsis* orthologue AT2G47820 encodes an arginine-glutamic acid dipeptide repeat protein, and PLAZA Homology enrichment analysis (https://bioinformatics.psb.ugent.be/plaza/) suggests that the molecular function of GO:0003677 (DNA binding) is associated with the product of the gene model.

The candidate gene MA_10433886g0010 associated with PC4 is linked with the GO-term GO:0007067 (mitosis) in the *P. abies* exAtlas (Nystedt et al., 2013) and its peak expression is found in flushing vegetative shoots although it is also expressed in, for example, buds. MA_19950g0010, the other candidate gene associated with PC4, appears to act in signal perception or transduction as its *Populus* orthologue and best BLASTX hits both encode WD-40 repeat family proteins (Table 3). This gene model has a contrasting expression pattern between early- and late-season buds (Figure 2). The gene models associated with PC3 have very low levels of expression in vegetative shoots that have just began to flush (Z4006TR02), while the gene models associated with PC4 and PC5 show relatively high expression in this tissue (Figure 2).

The presumed latent pathogen *S. polyspora*, the cherry spruce rust fungi (*T. areolata*) and the species causing needle cast diseases (*L. macrospora* and *R. kalkhoffii*; Horst, 2013; Kaitera, 2013; Pan et al., 2018; Rajala et al., 2013) met the criteria for targeted GWAS of specific, latent, pathogens. Trait-association mapping with presence/absence or relative abundances of *S. polyspora* or *R. kalkhoffii* yielded no QTLs, whereas two QTLs for *L. macrospora* presence/absence in the host phyllospere communities and one for *T. areolata* presence/absence were identified (Table 2). None of these QTLs was shared with the QTLs identified with the 1% most abundant OTUs; these traits were associated on independent SNPs, and appeared to be located at different positions in the genome (Table S4). The *L. macrospora* presence/absence in the host phyllosphere communities was associated with SNPs positioned in or adjacent to the Norway spruce gene models MA_97571g0020 and MA_10432519g0010 and the gene model MA_10g0010 was associated with *T. areolata* presence/absence (Table 3). No functional information is associated with MA_97571g0020 in the Norway spruce genome portal,
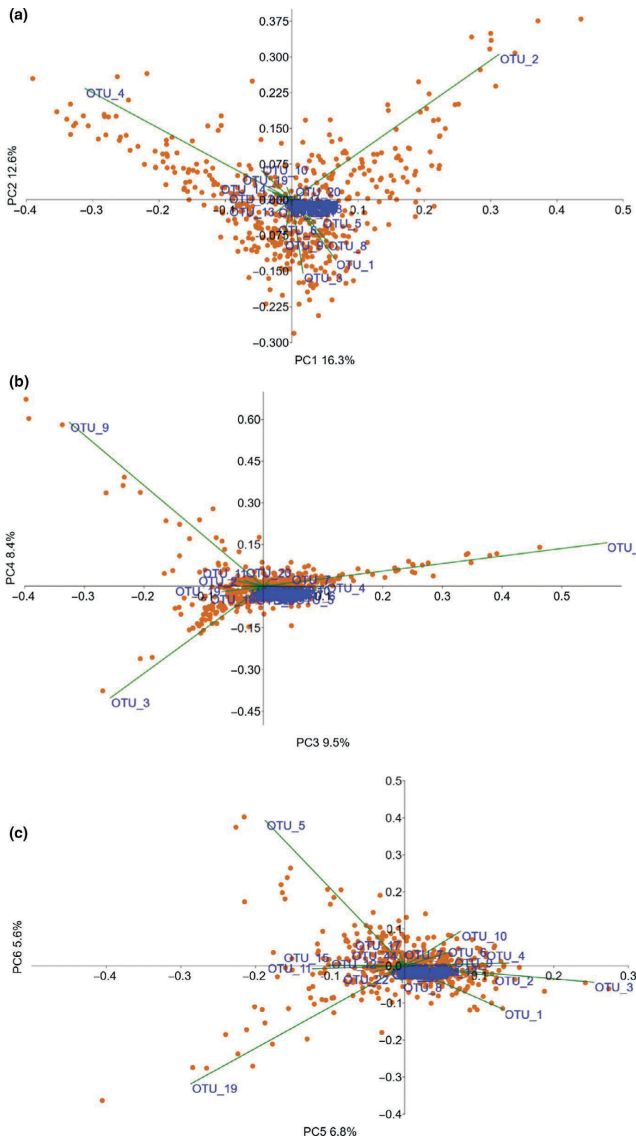
FIGURE 1 Principal component analysis of the operational taxonomic unit (OTU)/species abundance data. Biplots of the PCA on the abundance data of the 1% most heavily sequenced OTUs, used for the GWAS analysis. Orange symbols represent the individual samples. Green lines with blue OTU names represent the OTU loadings on the axis: (a) PC1 and PC2, (b) PC3 and PC4, and (c) PC5 and PC6

but the PFAM ID PF03801 (HEC/Ndc80p family) is associated with MA_10432519g0010; high relative expression levels of this gene model are found in young vegetative shoots (Z4006TR01) and in immature male cones in the *P. abies* exAtlas (Figure 2). MA_10g0010, associated with *T. areolata,* encodes a previously undescribed class III peroxidase. This candidate gene has a very distinct expression pattern, high expression levels of this gene model are found in young vegetative shoots, developing buds, wood and girdled twigs (Figure 2), and very high relative expression is found in immature female cones (Figure 2).

TABLE 2 Significant association in the GWA study

Trait[a]	SNP[b]	Allele	SNP feature[c]	Frequency[d]	PVE[e]
PC3	MA_24477_24501	T*C	downstream_gene	0.649	
	MA_31029_9337	C*G	missense	0.633	
	MA_10428833_21190	C*T	upstream_gene	0.552	
PC4	MA_19950_16139	G*C	upstream_gene	0.533	
	MA_10433886_12255	T*A	upstream_gene	0.682	1.2%
PC5	MA_208236_3389	A*G	stop_retained	0.565	
Lirula macrospora	MA_97571_20468	C*A	upstream_gene	0.566	
	MA_10432519_8378	C*A	synonymous	0.566	
Thekopsora areolata	MA_10_25927	G*A	upstream_gene	0.993	

[a]The trait upon which the marker associates, PC3–PC5 indicate the associations with loadings on the respective PC, and *L. macrospora* and *T. areolata* specify associations with the presence/absence data of these fungi among the samples.

[b]The SNP name consists of the contig (MA_number) and SNP position on the contig. For example, the first SNP MA_24477_24501 was located on contig MA_24477 at position 24,501 bp.

[c]Allelic variation associated with the SNP.

[d]Stability selection probability inclusion ratios for markers declared significant.

[e]Phenotypic variance explained (only values larger than 1.0% are displayed).

4 | DISCUSSION

Foliar fungi of conifers are commonly predicted to be horizontally transmitted and the phyllosphere community of a given tree would be expected to represent a sample of the fungi present in the environment (Millberg et al., 2015; Rodriguez et al., 2009; Terhonen et al., 2019). In angiosperm tree species, the genetic distance between host tree individuals may have a critical impact on assembly of the foliar fungal community (Ahlholm, Helander, Henriksson, Metzler, & Saikkonen, 2002; Cordier et al., 2012). The impact of the tree genotype is unclear in conifers where some studies report an effect of the host genotype on community composition (Rajala et al., 2013), and others find no such correlation (Eusemann et al., 2016). In this study, combining ITS sequencing and GWAS techniques, we provide further evidence that host genotype influences the composition of the phyllosphere community in the significant marker trait-association mapping of the variation in communities associated with dormant Norway spruce buds.

To our knowledge, no studies of conifer phyllosphere communities have used vegetative buds as a template. However, in a transcriptomic study of various Norway spruce tissues the highest frequency of fungal transcripts was found in needle and bud samples (Delhomme et al., 2015), suggesting that dormant vegetative buds, like needles, may be colonized by phyllosphere fungi. The results from our study further support this, as the taxonomic identities of many of the most heavily sequenced OTUs show affinity to previously reported members of the Norway spruce phyllosphere fungal community. For instance, *Phialophora sessilis*, *Taphrina carpini*, *Ceramothyrium* sp., *Phoma herbarum*, *Aureobasidium pullulans*, *Sydowia polyspora*, *Rhizosphaera kalkhoffii* and *Lirula macrospora* are species which have been reported from studies on Norway spruce needle communities (Menkis et al., 2015; Nguyen et al., 2016; Rajala et al., 2013, 2014). The frequencies and abundances of phyllosphere fungi differ depending on the species

present in the local environment (Nguyen et al., 2016; Rodriguez et al., 2009), physiological status of the host tree (Menkis et al., 2015; Rajala et al., 2014), tissue sampled and sample handling. The current study is an illustration of this, as the fraction of yeast-like fungi was higher here compared to earlier studies. The presence of a large fraction of yeast species may be due to the use of nonsurface-sterilized explants for DNA extraction, unlike in some previous studies (Nguyen et al., 2016; Rajala et al., 2013). This indicates a need for a similar marker trait association approach with both surface sterilized and nonsterilized explants and perhaps with the amplification and sequencing of the ITS2 region from several single buds from each tree instead of from a pooled sample. However, extraction and community sequencing of nonsurface-sterilized material as in our study, and the study by Menkis et al. (2015), indicates that the Norway spruce phyllosphere may have quite abundant, yet undescribed and possibly specific, epiphytes as illustrated by OTU_4. This unidentified taxon was one of the most common fungi in the phyllosphere community in the current study and the study by Menkis et al. (2015). It dominated on the first two PCA axes together with the basidiomycete yeast *Curvibasidium cygneicollum*, the common outdoor mould *Cladosporium herbarum*, to some extent the weak pathogen *S. polyspora*, and *Filobasidium wieringae*, suggesting that these axes may be driven primarily by highly frequent and highly abundant phylloplane fungi. It has previously been reported that tree genotypes can influence the occurrence of leaf epiphytes (Bálint et al., 2013; Cordier et al., 2012), but no marker-trait associations were detected with the two first PCs in this study. Possibly a much larger number of trees would be needed to pick up any association with the community on dormant buds. Other PCs were driven more strongly by well-known needle and wood colonizers such as *S. polyspora*, the biotrophic cherry spruce rust fungus *Thekopsora areolata* (primarily PC3 and PC4) and *L. macrospora* (PC5 and PC6).

Several, probable latent pathogens were identified among the most abundant species: *T. areolata*, *R. kalkhoffii*, *L. macrospora*,

**TABLE 3** Candidate Norway spruce gene models associated with the community composition and pathogen presence

Trait[a]	Candidate gene	GO terms[b]	Orthologues[c]	Best BlastX hit[d]
PC3	MA_24477g0010	GO:0005215—transporter activity, GO:0005525—GTP binding, GO:0005794—Golgi apparatus, GO:0006471—protein ADP-ribosylation, GO:0006886—intracellular protein transport, GO:0007264—small GTPase mediated signal transduction, GO:0016192—vesicle-mediated transport, GO:0016787—hydrolase activity	Potri.002G191400.1, AT1G10630.1	XP_024995498.1, ADP-ribosylation factor 2 isoform X1 [*Cynara cardunculus* var. *scolymus*]
	MA_31029g0010	N/A	Potri.014G151800.1, AT3G59340.1	XP_023873354.1, solute carrier family 35 member F2 [*Quercus suber*]
	MA_10428833g0010	N/A	N/A	ABR18169.1, unknown [*Picea sitchensis*]
	MA_10428833g0020	GO:0004842—ubiquitin-protein ligase activity, GO:0005634—nucleus, GO:0006817—phosphate ion transport, GO:0009626—plant-type hypersensitive response, GO:0009627—systemic acquired resistance, GO:0009697—salicylic acid biosynthetic process, GO:0009751—response to salicylic acid stimulus, GO:0010167—response to nitrate, GO:0010337—regulation of salicylic acid metabolic process, GO:0016036—cellular response to phosphate starvation, GO:0046872—metal ion binding, GO:0080021—response to benzoic acid stimulusGO:0042742—defence response to bacterium	Potri.016G064600.1, AT1G02860.1	ABR18169.1, unknown [*Picea sitchensis*]
PC4	MA_19950g0010	N/A	Potri.014G027200.1	XP_006826575.1, protein ROOT INITIATION DEFECTIVE 3 isoform X2 [*Amborella trichopoda*]
	MA_10433886g0010	GO:0005634—nucleus, GO:0005739—mitochondrion, GO:0005829—cytosol, GO:0007067—mitosis, GO:0009507—chloroplast	Potri.006G263300.2, AT5G47690.3	XP_003635522.2, PREDICTED: sister chromatid cohesion protein PDS5 homolog A-like [*Vitis vinifera*]
PC5	MA_208236g0010	N/A	Potri.008G210200.1, AT2G47820.2	XP_020082821, uncharacterized protein LOC109706424 isoform X2 [*Ananas comosus*]
Lirula macrospora	MA_97571g0020	N/A	N/A	XP_006350818, PREDICTED: acyl-coenzyme A thioesterase 13-like [*Solanum tuberosum*]
	MA_97571g0010	GO:0003863—3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring) activity, GO:0005739—mitochondrion, GO:0008152—metabolic process	Potri.005G185400.1, AT5G09300.2	XP_020524476.1, 2-oxoisovalerate dehydrogenase subunit alpha 2, mitochondrial isoform X2 [*Amborella trichopoda*]
	MA_10432519g0010	N/A	Potri.001G208700.1, AT3G54630.1	XP_021663255.1, kinetochore protein ndc80 [*Hevea brasiliensis*]
Thekopsora areolata	MA_10g0010	GO:0008152—metabolic process, GO:0016491—oxidoreductase activity, GO:0016772—transferase activity, transferring phosphorus-containing groups, GO:0046872—metal ion binding	Potri.007G074700.1, AT5G47000.1	XP_006847800.1, peroxidase 31 [*Amborella trichopoda*]

[a]The trait upon which the candidate gene associates; PC1–PC6 indicate the associations with loadings on the respective PC; and *L. macrospora* and *T. areolata* indicate associations with the presence/absence data of these fungi among the samples.

[b]GO terms associated with the gene model in the Norway spruce v1.0 genome portal.

[c]Angiosperm orthologues to the Norway spruce gene models reported in the Norway spruce v1.0 genome portal.

[d]The best BLASTX hit recovered when querying the NCBI nonredundant protein database with the Norway spruce gene model (E-value cut-off E < 1.0E-10).
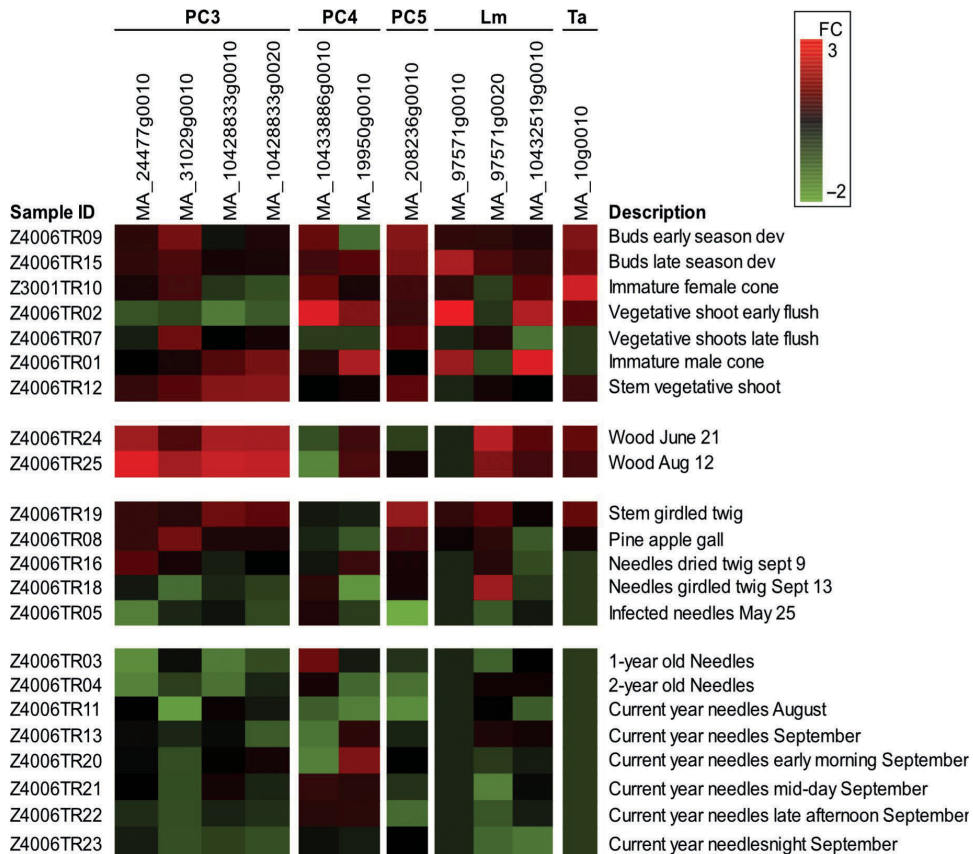
FIGURE 2 Expression of candidate genes in the *Picea abies* exAtlas. The heatmap depicts the relative expression levels (log$_2$ fold change [FC]) of the candidate genes in select samples collected from the *Picea abies* exAtlas (https://www.congenie.org). The heatmap shows high (upregulated) relative expression in red colours and low (downregulated) relative expression in green as indicated in the bar at the bottom

Botrytis cinerea, *P. herbarum* and *S. polyspora* (Hennon, 1990; Horst, 2013; Kaitera, 2013; Pan et al., 2018; Rajala et al., 2013). With the exception of *T. areolata*, these species have been reported from Norway spruce phyllosphere communities previously (Menkis et al., 2015; Nguyen et al., 2016; Rajala et al., 2013, 2014). The latent pathogens were, however, much more frequent in our study, found in up to 82% of the samples. It is possible that a surface sterilization treatment would have reduced the frequency of these taxa, but *S. polyspora*, *L. macrospora*, *T. areolata* and *R. kalkhoffii* were relatively highly abundant (File S3). Thus, our interpretation is that these four pathogens had colonized asymptomatic buds, acting as latent pathogens.

The interactions between trees and phyllosphere fungi are diverse and not fully understood. It has been reported that endophytic fungi are more diverse and abundant than pathogens within the phyllosphere community (Rodriguez et al., 2009; Terhonen et al., 2019), something that is reflected also in our study. Endophytic fungi in the phyllosphere have the potential to both enhance and reduce tree growth and fitness through various mechanisms (Rodriguez et al., 2009; Terhonen et al., 2019), one of them being the capacity of phyllosphere pathogens to act as both repressors and enablers of disease (Ridout & Newcombe, 2015). A recent study on *Populus* and *Melampsora* established that the interaction of *Melampsora* with endophytic disease repressors is governed by hierarchical contingency rules that determine when and where antagonists in the fungal community reduce plant disease severity, but that this interaction is forestalled by genetic variation for *Melampsora* resistance in *Populus* (Busby, Crutsinger,

Barbour, & Newcombe, 2019). If this is a general rule in the interaction between phyllosphere endophytes, pathogens and their hosts, identifying host genetic variation that restricts the colonization by pathogenic fungi or covariates with particular phyllosphere endophyte communities may help to increase our understanding of the interaction between trees and their phyllosphere fungal communities. In this study we were successful in identifying nine such QTLs, highlighting various potential mechanisms.

The gene models MA_10428833g0010 and MA_10428833g0020, associated with PC3, appear to either belong to the same gene as indicated by the shared best BLASTX hit, ABR18169.1, or possibly represent a tandem duplication of Ubiquitin E3 ligase genes. Angiosperm orthologues of ABR18169.1 have been shown to be linchpins in the responses to nitrate and phosphate starvation (Park, Seo, & Chua, 2014; Peng et al., 2008) and to regulate plant immune responses (Yaeno & Iba, 2008). The potential connection to nitrate and phosphate starvation responses makes this a highly interesting locus as the nutrient status of *Quercus macrocarpa* and *Pinus monticola* leaves is one of the factors that influence their phyllosphere fungal community (Jumpponen & Jones, 2010; Larkin, Hunt, & Ramsey, 2012). Furthermore, a recent study has shown that an impaired phosphate accumulation capacity in *Arabidopsis* led to a strong shift in the phyllosphere fungal community (Finkel et al., 2019).

Several of the potential orthologues of the Norway spruce candidate genes associated with PC3 and PC4 indicate that shoot development and morphogenesis are important factors in shaping genotype-specific phyllosphere communities in this study. The poplar orthologue of MA_19950g0010, Potri.014G027200.1, encodes a WD-40 repeat family protein with similarity to *ROOT INITIATION DEFECTIVE 3* (*RID3*). *RID3* has been functionally characterized and the *RID3* locus controls cell division and shoot apical meristem formation in *Arabidopsis* (Tamaki et al., 2009), acting as a negative regulator of key genes. In this context it is noteworthy that the expression of MA_19950g0010 is reported in the *P. abies* exAtlas to be very low during the early phases of Norway spruce bud development just after growth termination when the buds go through a phase of rapid needle initiation and changes in starch and tannin distribution patterns (Hejnowicz & Obarska, 1995). The sister chromatid cohesion proteins PDS5s, the *Arabidopsis* orthologues of MA_104333886g0010 which is associated with PC4, are involved in mitosis, and depletion of this protein lead to severe effects on development, among other traits (Pradillo et al., 2015). Notably, MA_104333886g0010 is most highly expressed in the early phases of bud flush, which is characterized by high mitotic activity and also by changes in both morphology and starch and tannin distribution (Hejnowicz & Obarska, 1995). Interestingly, the candidate genes associated with PC3 show a concomitant downregulation in vegetative shoots that have just began to flush (Z4006TR02), compared to earlier and later phases of shoot development; the genes are differentially expressed between samples taken from needles and stem, indicating that the cellular processes the candidate genes are associated with are not ubiquitous in the phyllosphere and that their absence could be important for

colonization success for certain fungal taxa. The observations associated with PC3 and PC4 suggest that bud flush, and the changes in distribution of energy and defensive metabolites that the tissue transitions through, may be a critical phase in the colonization of Norway spruce by the studied phyllosphere fungal community and their later occurrence in dormant buds. This is in line with the work by Eusemann et al. (2016) who postulated that tree phenology was one of the drivers between assembly of the phyllosphere community in their study. Phenology (timing of bud burst) is under very strong genetic control in Norway spruce, and it has a narrow sense heritability of approximately 0.8 (Hannerz, Sonesson, & Ekberg, 1999). The function and expression patterns of each of the identified Norway spruce candidate genes, and perhaps the role of phenology, in the interaction between phyllosphere fungi and Norway spruce, will need to be tested in future experiments.

One of the candidate genes associated with the presence/absence of *L. macrospora* in the communities, MA_10432519g0010, has similarity to the HEC/Ndc80p family proteins. These proteins are part of the kinetochore complex, which provides an attachment site for spindle fibres in the centromere of chromosomes, and thus are important for cell division (Shin, Jeong, Park, Kim, & Lee, 2018). Disruption of the kinetochore complex may affect, for example, morphogenesis (Du & Dawe, 2007; Lermontova et al., 2011; Shin et al., 2018). It may be difficult to reconcile a potential role of MA_10432519g0010 in development and morphogenesis with control of *L. macrospora* colonization as disease symptoms are commonly seen on second-year needles (Butin, 1995). However, it has been suggested that *L. macrospora* infects flushing tissues in the spring with a latency period of about 1 year until symptoms are visible (Hennon, 1990), which would connect the peak transcriptional activity of the candidate gene with the crucial infection phase.

MA_10g0010, which harbours an SNP associated with the presence of *T. areolata* in the phyllosphere fungal community is a previously undescribed member of the class III peroxidase family. The gene model shows an expression pattern strongly associated with differentiating and lignifying tissues in the *P. abies* exAtlas, particularly with young female cones. Clearly the candidate gene is active in tissues susceptible to the pathogen, and basidiospores of *T. areolata* are thought to infect Norway spruce cones and young shoots in the spring (Hietala, Solheim, & Fossdal, 2007; Kuporevich & Transhel, 1957). Thus, it is not far-fetched to imagine that MA_10g0010 may be associated with processes in these tissues that control their vulnerability to colonization, possibly through cell-wall enforcement (Elfstrand et al., 2001; Fagerstedt, Kukkola, Koistinen, Takahashi, & Marjamaa, 2010; Kärkönen, Warinowski, Teeri, Simola, & Fry, 2009; Marjamaa et al., 2006) or processes controlling the timing of bud break and bud flush; overexpression of the class III peroxidase *SPI2* in Norway spruce plants led to significant delays of these process compared to in wild-type Norway spruce plants (Clapham, Häggman, Elfstrand, Aronen, & Arnold, 2003; Elfstrand et al., 2001).

Combining molecular identification and barcoding of phyllosphere fungi with GWAS provides insight into host-encoded factors

affecting the assembly of phyllosphere communities and the colonization of needle pathogens. Taken together, GWAS results suggest that processes in the morphogenesis and flush of the Norway spruce shoot exert a strong influence on the dominant players in the phyllosphere community detected in dormant buds.

AUTHOR CONTRIBUTIONS

Conception of the study: Å.O., J.S., M.E., M.R.G.G., K.L., B.K. and H.X.W.; planning the study: M.E., Å.O., J.B., L.Z., M.R.G.G., K.L. and J.S.; execution of the study: M.E., J.B., L.Z. and M.R.G.G.; drafting the manuscript: M.E., J.B. and L.Z.; all authors read and commented on the text. M.E. wrote the final manuscript, and all authors read and approved the final version.

DATA AVAILABILITY STATEMENT

ITS7/ITS4 raw read data are stored at the NCBI Sequence Read Archive in the Bioproject PRJNA590990. Contact M.Rosario. Garcia@slu.se for genotype information

ORCID

Malin Elfstrand https://orcid.org/0000-0002-0214-5284
John Baison https://orcid.org/0000-0001-9919-907X
Åke Olson https://orcid.org/0000-0001-8998-6096
Karl Lundén https://orcid.org/0000-0002-0817-2419
Jan Stenlid https://orcid.org/0000-0002-5344-2094
M. Rosario García-Gil https://orcid.org/0000-0002-6834-6708

REFERENCES

Ahlholm, J. U., Helander, M., Henriksson, J., Metzler, M., & Saikkonen, K. (2002). Environmental conditions and host genotype direct genetic diversity of *Venturia ditricha*, a fungal endophyte of birch trees. *Evolution*, *56*, 1566–1573. https://doi.org/10.1111/j.0014-3820.2002.tb01468.x

Arnold, A. E., Henk, D. A., Eells, R. L., Lutzoni, F., & Vilgalys, R. (2007). Diversity and phylogenetic affinities of foliar fungal endophytes in loblolly pine inferred by culturing and environmental PCR. *Mycologia*, *99*, 185–206. https://doi.org/10.1080/15572536.2007.11832578

Baison, J., Vidalis, A., Zhou, L., Chen, Z.-Q., Li, Z., Sillanpää, M. J., … García-Gil, M. R. (2019). Genome-Wide Association Study (GWAS) identified novel candidate loci affecting wood formation in Norway spruce. *The Plant Journal*, *100*, 83–100. https://doi.org/10.1111/tpj.14429

Bálint, M., Tiffin, P., Hallström, B., O'Hara, R. B., Olson, M. S., Fankhauser, J. D., … Schmitt, I. (2013). Host genotype shapes the foliar fungal microbiome of balsam poplar (*Populus balsamifera*). *PLoS ONE*, *8*, e53987. https://doi.org/10.1371/journal.pone.0053987

Bernhardsson, C., Vidalis, A., Wang, X., Scofield, D. G., Schiffthaler, B., Baison, J., … Ingvarsson, P. K. (2019). An ultra-dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway spruce (*Picea abies*). *G3: Genes|genomes|genetics*, *9*, 1623–1632. https://doi.org/10.1534/g3.118.200840

Busby, P. E., Crutsinger, G., Barbour, M., & Newcombe, G. (2019). Contingency rules for pathogen competition and antagonism in a genetically based, plant defense hierarchy. *Ecology and Evolution*, *9*, 6860–6868. https://doi.org/10.1002/ece3.5253

Butin, H. H. (1995). *Tree diseases and disorders: Causes, biology, and control in forest and amenity trees*. Oxford: Oxford University Press.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*, 80–92. https://doi.org/10.4161/fly.19695

Clapham, D. H., Häggman, H., Elfstrand, M., Aronen, T., & von Arnold, S. (2003). Transformation of Norway spruce (*Picea abies*) by particle bombardment. In J. F. Jackson & H. F. Linskens (Eds.), *Genetic transformation of plants* (pp. 127–146). Berlin: Springer.

Clemmensen, K. E., Bahr, A., Ovaskainen, O., Dahlberg, A., Ekblad, A., Wallander, H., … Lindahl, B. D. (2013). Roots and associated fungi drive long-term carbon sequestration in boreal forest. *Science*, *339*, 1615–1618. https://doi.org/10.1126/science.1231923

Clemmensen, K. E., Ihrmark, K., Durling, M. B., & Lindahl, B. D. (2016). Sample preparation for fungal community analysis by high-throughput sequencing of barcode amplicons. In F. Martin & S. Uroz (Eds.), *Microbial environmental genomics (MEG)* (pp. 61–88). New York: Springer.

Cordier, T., Robin, C., Capdevielle, X., Desprez-Loustau, M.-L., & Vacher, C. (2012). Spatial variability of phyllosphere fungal assemblages: Genetic distance predominates over geographic distance in a European beech stand (*Fagus sylvatica*). *Fungal Ecology*, *5*, 509–520. https://doi.org/10.1016/j.funeco.2011.12.004

Delhomme, N., Sundström, G., Zamani, N., Lantz, H., Lin, Y.-C., Hvidsten, T. R., … Street, N. R. (2015). Serendipitous meta-transcriptomics: The fungal community of Norway spruce (*Picea abies*). *PLoS ONE*, *10*, e0139080. https://doi.org/10.1371/journal.pone.0139080

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*, 491–498. https://doi.org/10.1038/ng.806

Du, Y., & Dawe, R. K. (2007). Maize NDC80 is a constitutive feature of the central kinetochore. *Chromosome Research*, *15*, 767–775. https://doi.org/10.1007/s10577-007-1160-z

Elfstrand, M., Fossdal, C., Sitbon, F., Olsson, O., Lönneborg, A., & von Arnold, S. (2001). Overexpression of the endogenous peroxidase-like gene spi 2 in transgenic Norway spruce plants results in increased total peroxidase activity and reduced growth. *Plant Cell Reports*, *20*, 596–603. https://doi.org/10.1007/s002990100360

Eusemann, P., Schnittler, M., Nilsson, R. H., Jumpponen, A., Dahl, M. B., Würth, D. G., … Unterseher, M. (2016). Habitat conditions and phenological tree traits overrule the influence of tree genotype in the needle mycobiome–*Picea glauca* system at an arctic treeline ecotone. *New Phytologist*, *211*, 1221–1231. https://doi.org/10.1111/nph.13988

Fagerstedt, K. V., Kukkola, E. M., Koistinen, V. V. T., Takahashi, J., & Marjamaa, K. (2010). Cell wall lignin is polymerised by class iii secretable plant peroxidases in Norway spruce. *Journal of Integrative Plant Biology*, *52*, 186–194. https://doi.org/10.1111/j.1744-7909.2010.00928.x

Farjon, A., & Page, C. N. (1999). *Conifers. Status survey and conservation action plan*. Gland, Switzerland: IUCN.

Finkel, O. M., Salas-González, I., Castrillo, G., Spaepen, S., Law, T. F., Jones, C. D., & Dangl, J. L. (2019). The effects of soil phosphorous content on microbiota are driven by the plant phosphate starvation response. *bioRxiv*, 608133. https://doi.org/10.1101/608133

Gao, H., Wu, Y., Li, J., Li, H., Li, J., & Yang, R. (2014). Forward LASSO analysis for high-order interactions in genome-wide association study. *Briefings in Bioinformatics*, 15, 552–561. https://doi.org/10.1093/bib/bbt037

Hammer, O., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4, 1–9.

Hannerz, M., Sonesson, J., & Ekberg, I. (1999). Genetic correlations between growth and growth rhythm observed in a short-term test and performance in long-term field trials of Norway spruce. *Canadian Journal of Forest Research*, 29, 768–778. https://doi.org/10.1139/x99-056

Hanso, M., & Drenkhan, R. (2012). *Lophodermium* needle cast, insect defoliation and growth responses of young Scots pines in Estonia. *Forest Pathology*, 42, 124–135. https://doi.org/10.1111/j.1439-0329.2011.00728.x

Hejnowicz, A., & Obarska, E. (1995). Structure and development of vegetative buds, from the lower crown of *Picea abies*. *Annals of Forest Science*, 52, 433–447.

Hennon, P. E. (1990). Sporulation of *Lirula macrospora* and symptom development on Sitka spruce in southeast Alaska. *Plant Disease*, 74, 316–319. https://doi.org/10.1094/PD-74-0316

Hietala, A. M., Solheim, H., & Fossdal, C. G. (2007). Real-time PCR-based monitoring of DNA pools in the tri-trophic interaction between Norway spruce, the rust *Thekopsora areolata*, and an opportunistic ascomycetous *Phomopsis* sp. *Phytopathology*, 98, 51–58.

Horst, R. K. (2013). Needle casts. In R. K. Horst (Ed.), *Westcott's plant disease handbook* (pp. 245–249). Dordrecht: Springer.

Horton, M. W., Bodenhausen, N., Beilsmith, K., Meng, D., Muegge, B. D., Subramanian, S., … Bergelson, J. (2014). Genome-wide association study of *Arabidopsis thaliana* leaf microbial community. *Nature Communications*, 5, 5320. https://doi.org/10.1038/ncomms6320

Ihrmark, K., Bödeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., … Lindahl, B. D. (2012). New primers to amplify the fungal ITS2 region-evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology*, 82, 666–677. https://doi.org/10.1111/j.1574-6941.2012.01437.x

Jumpponen, A., & Jones, K. L. (2010). Seasonally dynamic fungal communities in the *Quercus macrocarpa* phyllosphere differ between urban and nonurban environments. *New Phytologist*, 186, 496–513. https://doi.org/10.1111/j.1469-8137.2010.03197.x

Kaitera, J. (2013). *Thekopsora* and *Chrysomyxa* cone rusts damage Norway spruce cones after a good cone crop in Finland. *Scandinavian Journal of Forest Research*, 28, 217–222.

Kärkönen, A., Warinowski, T., Teeri, T. H., Simola, L. K., & Fry, S. C. (2009). On the mechanism of apoplastic H_2O_2 production during lignin formation and elicitation in cultured spruce cells—Peroxidases after elicitation. *Planta*, 230, 553–567. https://doi.org/10.1007/s00425-009-0968-5

Kubartová, A., Ottosson, E., Dahlberg, A., & Stenlid, J. (2012). Patterns of fungal communities among and within decaying logs, revealed by 454 sequencing. *Molecular Ecology*, 21, 4514–4532. https://doi.org/10.1111/j.1365-294X.2012.05723.x

Kuporevich, V., & Transhel, V. (1957). *Cryptogamic plants of the USSR: Rust fungi*. Jerusalem: Keter Press.

Larkin, B. G., Hunt, L. S., & Ramsey, P. W. (2012). Foliar nutrients shape fungal endophyte communities in Western white pine (*Pinus monticola*) with implications for white-tailed deer herbivory. *Fungal Ecology*, 5, 252–260. https://doi.org/10.1016/j.funeco.2011.11.002

Lermontova, I., Koroleva, O., Rutten, T., Fuchs, J., Schubert, V., Moraes, I., … Schubert, I. (2011). Knockdown of CENH3 in *Arabidopsis* reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. *The Plant Journal*, 68, 40–50. https://doi.org/10.1111/j.1365-313X.2011.04664.x

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26, 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, Z., Hallingbäck, H. R., Abrahamsson, S., Fries, A., Andersson Gull, B., Sillanpää, M. J., & García-Gil, M. R. (2014). Functional multi-locus QTL mapping of temporal trends in Scots pine wood traits. *G3: Genes|Genomes|Genetics*, 4(12), 2365–2379. https://doi.org/10.1534/g3.114.014068

Li, Z., & Sillanpää, M. J. (2015). Dynamic quantitative trait locus analysis of plant phenomic data. *Trends in Plant Science*, 20, 822–833. https://doi.org/10.1016/j.tplants.2015.08.012

Marjamaa, K., Hildén, K., Kukkola, E., Lehtonen, M., Holkeri, H., Haapaniemi, P., … Lundell, T. (2006). Cloning, characterization and localization of three novel class III peroxidases in lignifying xylem of Norway spruce (*Picea abies*). *Plant Molecular Biology*, 61, 719–732. https://doi.org/10.1007/s11103-006-0043-6

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20, 1297–1303. https://doi.org/10.1101/gr.107524.110

Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, 417–473. https://doi.org/10.1111/j.1467-9868.2010.00740.x

Menkis, A., Marciulynas, A., Gedminas, A., Lynikiene, J., & Povilaitiene, A. (2015). High-throughput sequencing reveals drastic changes in fungal communities in the phyllosphere of Norway spruce (*Picea abies*) following invasion of the spruce bud scale (*Physokermes piceae*). *Microbial Ecology*, 70, 904–911. https://doi.org/10.1007/s00248-015-0638-z

Millberg, H., Boberg, J., & Stenlid, J. (2015). Changes in fungal community of Scots pine (*Pinus sylvestris*) needles along a latitudinal gradient in Sweden. *Fungal Ecology*, 17, 126–139. https://doi.org/10.1016/j.funeco.2015.05.012

Moler, E. R. V., & Aho, K. (2018). Whitebark pine foliar fungal endophyte communities in the southern Cascade Range, USA: Host mycobiomes and white pine blister rust. *Fungal Ecology*, 33, 104–114. https://doi.org/10.1016/j.funeco.2018.02.003

Nguyen, D., Boberg, J., Ihrmark, K., Stenström, E., & Stenlid, J. (2016). Do foliar fungal communities of Norway spruce shift along a tree species diversity gradient in mature European forests? *Fungal Ecology*, 23, 97–108. https://doi.org/10.1016/j.funeco.2016.07.003

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., … Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497, 579–584. https://doi.org/10.1038/nature12211

Pan, Y., Ye, H., Lu, J., Chen, P., Zhou, X.-.-D., Qiao, M., … Yu, Z.-.-F. (2018). Isolation and identification of *Sydowia polyspora* and its pathogenicity on *Pinus yunnanensis* in Southwestern China. *Journal of Phytopathology*, 166, 386–395.

Park, B. S., Seo, J. S., & Chua, N.-H. (2014). NITROGEN LIMITATION ADAPTATION Recruits PHOSPHATE2 to target the phosphate transporter PT2 for degradation during the regulation of *Arabidopsis* phosphate homeostasis. *The Plant Cell*, 26, 454–464.

Peng, M., Hudson, D., Schofield, A., Tsao, R., Yang, R., Gu, H., … Rothstein, S. J. (2008). Adaptation of *Arabidopsis* to nitrogen limitation involves induction of anthocyanin synthesis which is controlled by the *NLA* gene. *Journal of Experimental Botany*, 59, 2933–2944. https://doi.org/10.1093/jxb/ern148

Pradillo, M., Knoll, A., Oliver, C., Varas, J., Corredor, E., Puchta, H., & Santos, J. L. (2015). Involvement of the cohesin cofactor PDS5

(SPO76) during meiosis and DNA repair in *Arabidopsis thaliana*. *Frontiers Plant Science*, 6. https://doi.org/10.3389/fpls.2015.01034

Prunier, J., Verta, J.-P., & MacKay, J. J. (2016). Conifer genomics and adaptation: At the crossroads of genetic diversity and genome function. *New Phytologist*, 209, 44–62. https://doi.org/10.1111/nph.13565

Rajala, T., Velmala, S. M., Tuomivirta, T., Haapanen, M., Müller, M., & Pennanen, T. (2013). Endophyte communities vary in the needles of Norway spruce clones. *Fungal Biology*, 117, 182–190. https://doi.org/10.1016/j.funbio.2013.01.006

Rajala, T., Velmala, S. M., Vesala, R., Smolander, A., & Pennanen, T. (2014). The community of needle endophytes reflects the current physiological state of Norway spruce. *Fungal Biology*, 118, 309–315. https://doi.org/10.1016/j.funbio.2014.01.002

Ridout, M., & Newcombe, G. (2015). The frequency of modification of *Dothistroma* pine needle blight severity by fungi within the native range. *Forest Ecology and Management*, 337, 153–160. https://doi.org/10.1016/j.foreco.2014.11.010

Rodriguez, R. J., White, J. F. Jr, Arnold, A. E., & Redman, R. S. (2009). Fungal endophytes: Diversity and functional roles. *New Phytologist*, 182, 314–330. https://doi.org/10.1111/j.1469-8137.2009.02773.x

Roman-Reyna, V., Pinili, D., Borjaa, F. N., Quibod, I., Groen, S. C., Mulyaningsih, E. S., ... Oliva, R. (2019). The rice leaf microbiome has a conserved community structure controlled by complex host-microbe interactions. *bioRxiv*, 615278

Rosling, A., Cox, F., Cruz-Martinez, K., Ihrmark, K., Grelet, G.-A., Lindahl, B. D., ... James, T. Y. (2011). Archaeorhizomycetes: Unearthing an ancient class of ubiquitous soil fungi. *Science*, 333, 876–879. https://doi.org/10.1126/science.1206958

Saikkonen, K. (2007). Forest structure and fungal endophytes. *Fungal Biology Reviews*, 21, 67–74. https://doi.org/10.1016/j.fbr.2007.05.001

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics*, 38, 595–619. https://doi.org/10.1146/annurev.ecolsys.38.091206.095646

Seena, S., & Monroy, S. (2016). Preliminary insights into the evolutionary relationships of aquatic hyphomycetes and endophytic fungi. *Fungal Ecology*, 19, 128–134. https://doi.org/10.1016/j.funeco.2015.07.007

Shin, J., Jeong, G., Park, J.-Y., Kim, H., & Lee, I. (2018). *MUN* (*MERISTEM UNSTRUCTURED*), encoding a SPC24 homolog of NDC80 kinetochore complex, affects development through cell division in *Arabidopsis thaliana*. *The Plant Journal*, 93, 977–991.

Sieber, T. N. (2007). Endophytic fungi in forest trees: Are they mutualists? *Fungal Biology Reviews*, 21, 75–89. https://doi.org/10.1016/j.fbr.2007.05.004

Tamaki, H., Konishi, M., Daimon, Y., Aida, M., Tasaka, M., & Sugiyama, M. (2009). Identification of novel meristem factors involved in shoot regeneration through the analysis of temperature-sensitive mutants of *Arabidopsis*. *The Plant Journal*, 57, 1027–1039. https://doi.org/10.1111/j.1365-313X.2008.03750.x

Team, R. (2015). *RStudio: integrated development for R.* Boston: RStudio Inc. Retrieved from http://www.rstudio.com

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., ... Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science*, 346(6213), 1256688–https://doi.org/10.1126/science.1256688

Terhonen, E., Blumenstein, K., Kovalchuk, A., & Asiegbu, F. O. (2019). Forest tree microbiomes and associated fungal endophytes: Functional roles and impact on forest health. *Forests*, 10, 42. https://doi.org/10.3390/f10010042

Vidalis, A., Scofield, D. G., Neves, L. G., Bernhardsson, C., García-Gil, M. R., & Ingvarsson, P. K. (2018). Design and evaluation of a large sequence-capture probe set and associated SNPs for diploid and haploid samples of Norway spruce (*Picea abies*). *bioRxiv*, https://doi.org/10.1101/291716

Voříšková, J., & Baldrian, P. (2012). Fungal community on decomposing leaf litter undergoes rapid successional changes. *The ISME Journal*, 7, 477. https://doi.org/10.1038/ismej.2012.116

Wallace, J. G., Kremling, K. A., Kovar, L. L., & Buckler, E. S. (2018). Quantitative genetics of the maize leaf microbiome. *Phytobiomes Journal*, 2, 208–224. https://doi.org/10.1094/PBIOMES-02-18-0008-R

Yaeno, T., & Iba, K. (2008). BAH1/NLA, a RING-type ubiquitin E3 Ligase, regulates the accumulation of salicylic acid and immune responses to *Pseudomonas syringae* DC3000. *Plant Physiology*, 148, 1032–1041.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.-

V

RESEARCH ARTICLE   **Open Access**

Effect of number of annual rings and tree ages on genomic predictive ability for solid wood properties of Norway spruce

Linghua Zhou[1], Zhiqiang Chen[1], Lars Olsson[2], Thomas Grahn[2], Bo Karlsson[3], Harry X. Wu[1,4,5], Sven-Olof Lundqvist[2,6†] and María Rosario García-Gil[1*†]

Abstract

Background: Genomic selection (GS) or genomic prediction is considered as a promising approach to accelerate tree breeding and increase genetic gain by shortening breeding cycle, but the efforts to develop routines for operational breeding are so far limited. We investigated the predictive ability (PA) of GS based on 484 progeny trees from 62 half-sib families in Norway spruce (*Picea abies* (L.) Karst.) for wood density, modulus of elasticity (MOE) and microfibril angle (MFA) measured with SilviScan, as well as for measurements on standing trees by Pilodyn and Hitman instruments.

Results: GS predictive abilities were comparable with those based on pedigree-based prediction. Marker-based PAs were generally 25–30% higher for traits density, MFA and MOE measured with SilviScan than for their respective standing tree-based method which measured with Pilodyn and Hitman. Prediction accuracy (PC) of the standing tree-based methods were similar or even higher than increment core-based method. 78–95% of the maximal PAs of density, MFA and MOE obtained from coring to the pith at high age were reached by using data possible to obtain by drilling 3–5 rings towards the pith at tree age 10–12.

Conclusions: This study indicates standing tree-based measurements is a cost-effective alternative method for GS. PA of GS methods were comparable with those pedigree-based prediction. The highest PAs were reached with at least 80–90% of the dataset used as training set. Selection for trait density could be conducted at an earlier age than for MFA and MOE. Operational breeding can also be optimized by training the model at an earlier age or using 3 to 5 outermost rings at tree age 10 to 12 years, thereby shortening the cycle and reducing the impact on the tree.

Background

Norway spruce is one of the most important conifer species in Europe in relation to economic and ecological aspects [1]. Breeding of Norway spruce started in the 1940s with phenotypic selection of plus-trees, first in natural populations and later in even-aged plantations [2]. Norway spruce breeding cycle is approximately 25–30 years long,

of which the production of seeds and the evaluation of the trees take roughly one-half of that time [3].

Genomic prediction using genome-wide dense markers or genomic selection (GS) was first introduced by Meuwissen [4]. The method modelling the effect of large numbers of DNA markers covering the entire genome and subsequently predict the genomic value of individuals that have been genotyped, but not phenotyped. As compared to the phenotypic mass selection based on a pedigree-based relationship matrix (**A** matrix), genomic prediction relies on constructing a marker-based relationship matrix (**G** matrix). The superiority of the G-

* Correspondence: m.rosario.garcia@slu.se
Sven-Olof Lundqvist and María Rosario García-Gil Shared last authorship
[1]Department of Forest Genetics and Plant physiology, Umeå Plant Science Centre, Swedish University of Agricultural Sciences, SE-901 83 Umeå, Sweden
Full list of author information is available at the end of the article

matrix is the result of a more precise estimation of genetic similarity based on Mendelian segregation that not only captures recent pedigree but also the historical pedigree [5–7], and corrects possible errors in the pedigree [8, 9].

There are multiple factors affecting genomic prediction accuracy such as the extent of linkage disequilibrium (LD) between the marker loci and the quantitative trait loci (QTL), which is determined by the density of markers and the effective population size (N_e). Increased accuracy with higher marker density has been reported in simulation [10] and empirical studies in multiple forest tree species including Norway spruce [11–14], and SNP position showed no significant effect [15–17]. Simulation [10] and empirical [18] studies also agree on the need of a high marker density in populations with larger effective size (N_e) in order to cover more QTLs under low LD in contributing to the phenotypic variance.

In forest tree species the accuracy of the genomic prediction model has been mainly tested in cross-validation designs where full-sibs and/or half-sibs progenies within a single generation are subdivided into training and validation sets [10, 19–22]. Model accuracy was reported to increase with larger training to validation set ratios [11, 17, 23], while the level of relatedness between the two sets is considered as a major factor [10, 15–17, 19, 24]. When genomic prediction is conducted across environments, the level of genotype by environment interaction (GxE) of the trait determines its efficiency [11, 20, 21, 25]. The number of families and progeny size have also been shown to affect model accuracy [11, 15].

As compared to the previously described factors, trait heritability and specially trait genetic architecture are intrinsic characteristics to the studied trait in a given population. Those two factors can also be addressed by choosing an adequate statistical model depending on the expected distribution of the marker effects [26]. Despite theory and some results indicate that complex genetic structures obtain better fit with models that assume equal contribution of all markers to the observed variation, traits like disease-resistance are better predicted with methods where markers are assumed to have different variances [13, 20, 22, 27, 28]. However, results in forestry so far indicate that statistical models have little impact on the GS efficiency [12, 17, 29].

In this study, we conducted a genomic prediction study for solid wood properties based on data from 23-year old trees from open-pollinated (OP) families of Norway spruce. We focused on wood density, microfibril angle (MFA) and modulus of elasticity/wood stiffness (MOE) measured both with SilviScan in the lab, on standing trees of Pilodyn penetration depth and Hitman velocity of sound. The measurement methods are detailed in the next section.

The specific aims of the study were: (i) to compare narrow-sense heritability (h^2) estimation, predictive ability (PA) and prediction accuracy (PC) of the pedigree-based (ABLUP) models with marker-based models based on data from measurements with SilviScan on increment cores and from Pilodyn and Hitman measurements on standing trees, (ii) to examine the effects on model PA and PC of different training-to-validation set ratios and different statistical methods, (iii) to compare some practical alternatives to implement early training of genomic prediction model into operational breeding.

Result
Narrow-sense heritability (h^2) of the phenotypic traits, predictive ability (PA) and predictive accuracy (PC) based on pedigree and maker data
In Table 1, narrow sense heritabilities (h^2) and Prediction Abilities (PA) based on ABLUP and GBLUP are compared for density, MFA and MOE based on cross-sectional averages at age 19 years, and for Pilodyn, Velocity and MOE_{ind} based on measurements with the bark at age 22 and 24 years, respectively. For density, MOE and Pilodyn, h^2 did not differ significantly between estimates based on the pedigree (ABLUP) and marker-based (GBLUP) methods taking standard error into account. For MFA, the pedigree-based h^2 was lower than the GBLUP estimate while for Velocity and MOE_{ind}, the pedigree-based h^2 was higher.

When using pedigree, the order of the traits by h^2 agrees with their order by PA estimates. Traits with higher h^2 tended to show also high PA estimates irrespective of the method. The ABLUP PA estimates were similar to the GBLUP estimates for density and Pilodyn, while for the rest of the traits GBLUP delivered slightly higher PA estimates, and significantly higher for MFA. The relative performances of ABLUP compared to GBLUP differed for MOE, Velocity and MOE_{ind}. The h^2 estimates for MOE were similar for both methods, while the PA estimate was higher for GBLUP. In the case of Velocity and MOE_{ind}, a higher h^2 based on pedigree contrasted with a slightly higher PA estimates based on marker data. Standardization of the PAs with the h values did not change the conclusions on the relative efficiencies of pedigree versus marker data-based estimates.

Marker-based PA and PC between increment core-based and standing-base wood quality traits
The marker-based PAs were generally 25–30% higher for traits density, MFA and MOE measured with SilviScan than for their respective standing tree-based method which measured with Pilodyn and Hitman. Concordantly, the h^2 values were 46, 65 and 55% higher based on Silviscan methods, respectively. However, if we compare PC of the increment core- and standing tree-based methods, they were similar, and PC of MOE_{ind} was even higher than that for MOE using GBLUP.

Zhou *et al. BMC Genomics*   (2020) 21:323

Page 3 of 12

Table 1 Trait heritability, predictive ability (PA) and predictive accuracy (PC) Predictive accuracy (PC) for density, MFA and MOE cross-sectional averages at tree age 19 years, for their proxies on the stems without removing the bark at tree ages 21 and 22 years. Standard errors are shown in within parenthesis

Trait	Narrow-sense heritability (standard error) (h^2)		Predictive ability (standard error) (PA)		Predictive Accuracy (PA/h)	
	ABLUP	GBLUP	ABLUP	GBLUP	ABLUP	GBLUP
density	0.70 (0.18)	0.69 (0.15)	0.30 (0.01)	0.29 (0.03)	0.36	0.35
MFA	0.04 (0.08)	0.17 (0.13)	0.04 (0.01)	0.16 (0.02)	0.20	0.39
MOE	0.27 (0.14)	0.31 (0.15)	0.15 (0.01)	0.22 (0.02)	0.29	0.39
Pilodyn	0.35 (0.15)	0.32 (0.14)	0.22 (0.01)	0.20 (0.01)	0.37	0.35
Velocity	0.16 (0.12)	0.11 (0.10)	0.10 (0.01)	0.13 (0.01)	0.25	0.39
MOEind	0.31(0.14)	0.17 (0.13)	0.17 (0.01)	0.19 (0.01)	0.31	0.46

ABLUP pedigree-based Best Linear Unbiased Predictor (BLUP); *GBLUP* genomic-based BLUP

Effects on PAs of the GS models ratios between the training and validation sets, and from the statistical method used

Figure 1 shows how the PA estimates change with increasing percentage of data used for training of the GS model (training set), and as a consequence decreasing validation set, on use of the five studied statistical methods: one based on pedigree data and four on marker information. For most of the traits, PA estimates showed a moderate increase with increasing training set, irrespective of the statistical method. Exceptions were observed for MFA and MOE with less clear trends and
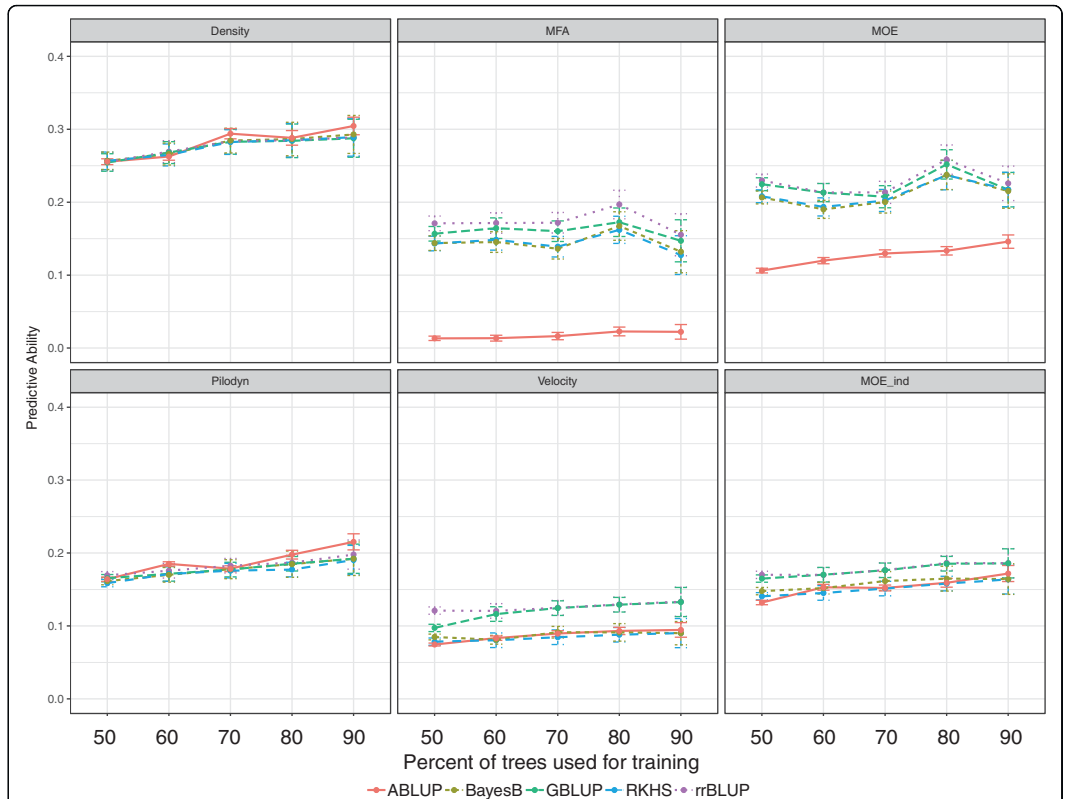
Fig. 1 Predictive ability obtained with different ratios of training set and validation set, using different statistical methods

the highest PA estimates at 80% of the trees in the training set. Figure 1 also shows that the PAs were consistently about 25–30% higher for density, MFA and MOE compared to their proxies-based om measurements with Pilodyn and Hitman: approximately 0.28 versus 0.18, 0.17 versus 0.13 and 0.25 versus 0.18, respectively.

For density and Pilodyn, all five methods resulted in very similar PA estimates across the ratios, while rrBLUP and GBLUP seemed superior for the rest of the traits, and mostly so for Velocity and MOE (Fig. 1). The rest of the analysis were conducted based on the GBLUP modelling method.

PAs on estimation of traits at reference age with models trained on data available at earlier ages

Figure 2 shows how well the cross-sectional averages of the different traits at the reference age 19 years were predicted by models trained based on data from the rings between pith and bark at increasing ages, using the GBLUP method. The calculations were performed with two representations of age: 1) Tree age counted from the establishment of the trial (calendar age) and 2) cambial age (ring number). In a plantation, the tree age of a planted tree is normally known but not the cambial age at breast height, as it depends on when the tree reached the breast height. For the trees originally accessed, almost 6000 trees from the two trials, this age ranged from tree age 2 to 15 years [30]. Among the 484 trees investigated in the current study, only 60 trees representing 33 families had reached breast height at tree age 3 years, 248 trees at 4 years and 410 at age 5 years (Fig. 2). This means that for tree age, data are only available from year 3, and then for only 12% of the trees. Those trees being identified based on fast longitudinal growth but also typically fast-growing radially. It was previously described a positive correlation of $R^2 = 0.67$ familywise between radial and height grown across almost 6000 trees [30]. Thereafter, the number of trees increased and reached the full number some years later. When studying the trees based on cambial age, the pattern is adverse with data for all trees at ring 1 but decreasing numbers when approaching the tree age of sampling. The number of trees included in this work at each tree and cambial age are shown with grey bars in Fig. 2.

For density, the estimated PAs showed a rising trend within a span of about 0.25–0.30 for the models based on both age types, after the first years. But the year-to-year fluctuations were more intense for models based on data organized on tree age. As MFA typically develops from high values at the lowest cambial ages via a rapid decrease to lower and more stable values from cambial age 8–12 years and on, one may expect that models trained on data from only low ages would have difficulties to predict properties at age 19 years. This was also
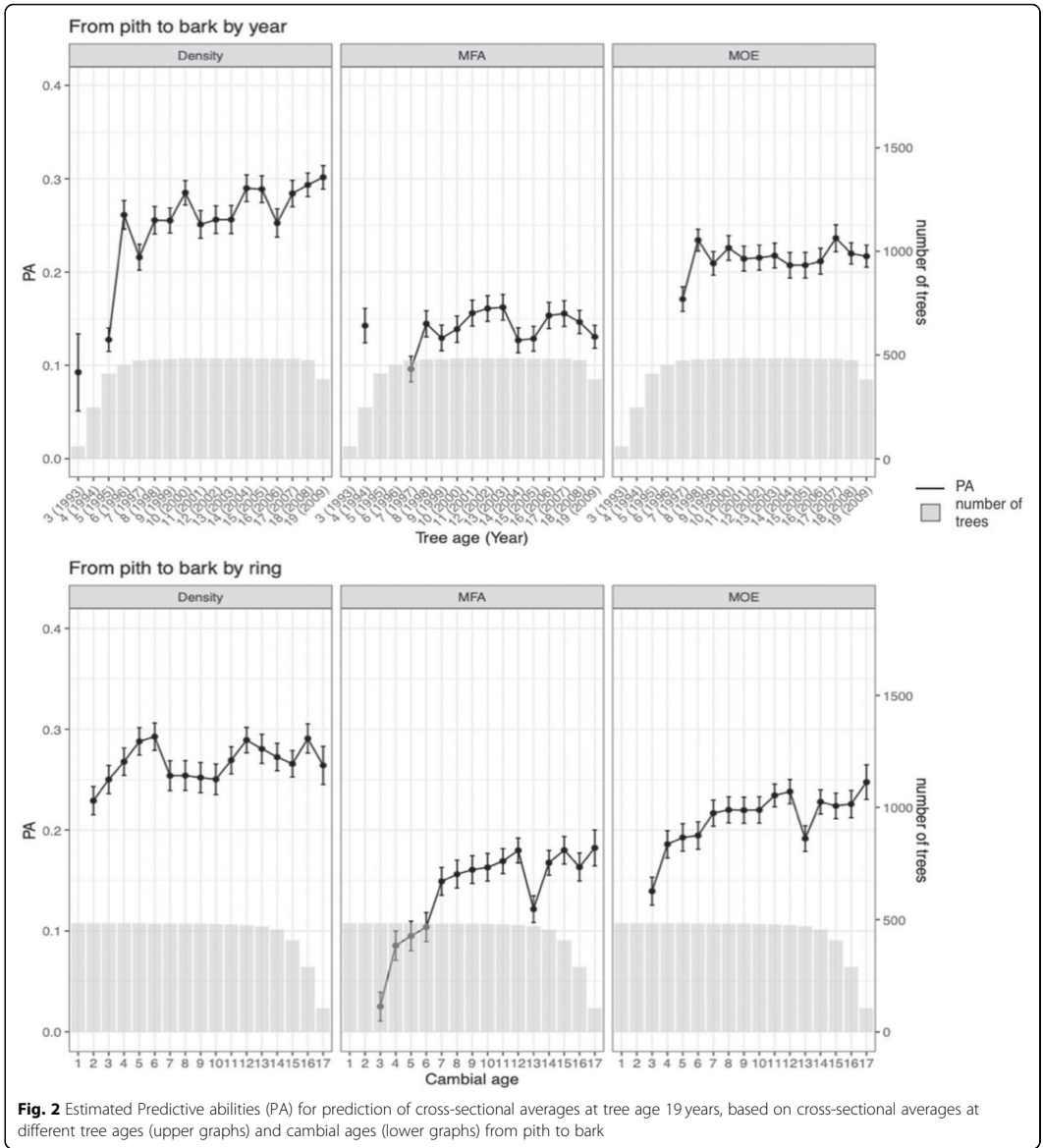
confirmed. We even obtained some negative PA values at early ages, such as years 1995 and 1996, and the PAs for cambial age-based models started from very low values, then increasing. The curves for MOE showed PAs developing at values in between those for density and MFA. This is logical, as MOE is influenced by both density and MFA, with particularly negative effects from the high MFAs at low cambial ages. At cambial age 13, MFA and MOE showed a drop in the cambial age-based PA estimates. Generally, the Figure indicates that genomic selection for density could be conducted at an earlier age than for MFA and MOE.

Search for optimal sampling and data for training of GS prediction models

Figure 2 showed estimated PAs of models trained on data from sampling different years, using data from all rings available at that age (except for the innermost ring). In this section instead of estimating PAs with the whole increment core from bark to pith, we estimated PAs with partial cores with different shorter depths to reduce the injury to the tree, as showed in Fig. 3a-d. This analysis was preformed based on tree age data only, as the cambial age of a ring can only be precisely known if the core is drilled to the pith which allowing all rings to be counted.

Each row of the figures represents a tree age when cores are samples, starting at age 3 years when the first 60 trees formed a ring at breast height, ending at the bottom with the reference age 19 years with17 rings. Each column represents a depth of coring, counted in numbers of rings. As one more ring is added each year, thus also to the maximum possible depth on coring, the tables are diagonal. The uppermost diagonal represents models trained on data from the 60 (12%) trees which had reached breast height at age 3. The diagonal next below represents models based on the 243 (51%) trees with rings at age 4, etc. The PAs shown below the three uppermost diagonals represent models trained of data from more than 90% of the trees. The PAs were calculated from the cross-validation, based on data from the trees on which the respective models were trained. This means that the PAs of the three uppermost diagonals are based only on fast-growing trees not fully representative for the trials. Many of the highest PAs found occur along these diagonals. Due to their trees' special growth, only PAs based on more than 90% of the trees will be further commented.

For wood density, Fig. 3b, the variations in predictability show an expected general pattern: The PAs increased with the increase of tree age on coring, and also with the increase of depth, the increase of number of rings from which the cross-sectional averages were calculated and exploited on training of the prediction models. The

Zhou *et al. BMC Genomics*   (2020) 21:323

Page 5 of 12

Fig. 2 Estimated Predictive abilities (PA) for prediction of cross-sectional averages at tree age 19 years, based on cross-sectional averages at different tree ages (upper graphs) and cambial ages (lower graphs) from pith to bark
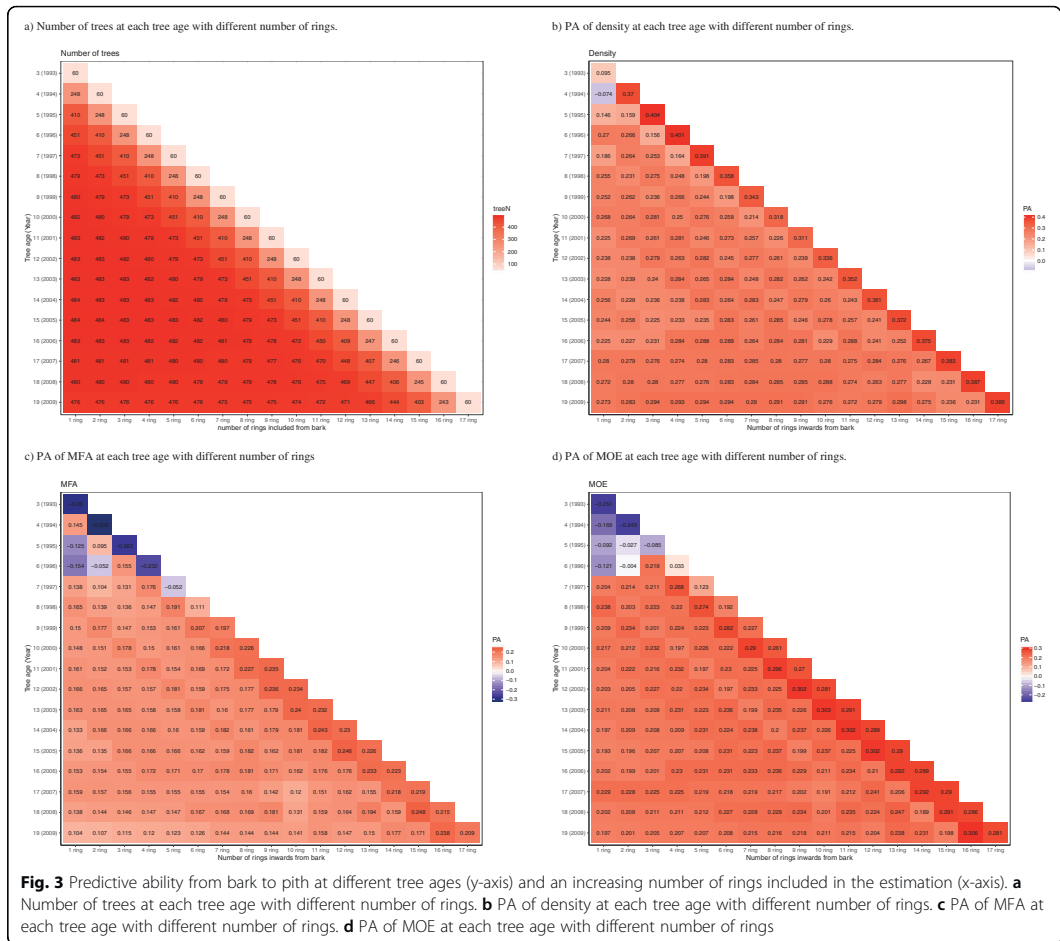
highest values, 0.29, are obtained at age 19 years, but then also data from the reference year are included on training the prediction model. An example of quite high PAs at lower ages and depths: For coring at tree ages 10–12 years and using data from the 3–5 outermost rings, all alternatives gave PA values of 0.26–0.29.

For MFA, a trait with low heritability, the PA values are low as already shown in Fig. 2 and the pattern in Fig.

3c is not easy to interpret. Here, the same set of alternatives of samples at tree ages 10–12 and depths 3–5 outermost rings gave PA values of 0.15–0.18, compared to the maximum of 0.19 among all alternatives using 90% of the trees. The values are lower at the highest ages. Streaks of higher and lower values can be imagined along the diagonals. The pattern for MOE in Fig. 3d is similar to that of MFA, but on higher level. Training on

Fig. 3 Predictive ability from bark to pith at different tree ages (y-axis) and an increasing number of rings included in the estimation (x-axis). **a** Number of trees at each tree age with different number of rings. **b** PA of density at each tree age with different number of rings. **c** PA of MFA at each tree age with different number of rings. **d** PA of MOE at each tree age with different number of rings

data from coring at ages and to depths as above gave PA values of 0.20–0.23, compared to the corresponding maximum of 0.25.

Discussion

We have conducted a genomic prediction study for solid wood properties assessed on increment cores from Norway spruce trees with SilviScan derived data from pith to bark, using properties of annual rings formed up to tree age 19 years as the reference age.

On Norway spruce operational breeding, the use of OP families is preferable because it does not require expensive control crosses. The only action required is to collect cones where progenies are typically assumed to be half-sibs. Thus, OP families permit the evaluation of large numbers of trees at lower costs and efforts than structured crossing designs. We investigated narrow-

sense heritability estimation with ABLUP and marker-based GBLUP and the effect on PA from using different training-to-validation set ratios, as well as different statistical methods. Further, we investigated what level of precision can be reached when training the models with data from trees at different ages, and 5also compared results for the solid wood properties with those for their proxies. We also estimated the level of PAs reached when coring to different depths from the bark at different tree ages. The motivation was to find cost-effective methods for GS with minimum impact on the trees during the acquisition of data for training the prediction models.

Narrow-sense heritability (h^2)

In our study, PA estimates for both pedigree and marker-based methods were consistent with their

Zhou *et al. BMC Genomics*   (2020) 21:323

Page 7 of 12

respective h^2 estimates. A conifer literature review indicates that the level of consistency varies across studies [8, 18–20]. In our study, h^2 estimation of density, MOE and Pilodyn were similar for ABLUP and GBLUP; for Velocity and MOE_{ind}, ABLUP had higher h^2 estimation and for MFA, GBLUP achieved higher h^2 estimation. In a previous study conducted on full-sib progenies in Norway spruce, however, the ABLUP-based h^2 were reported higher in all three standing-tree-based measurements [11]. Instead, other conifer studies based on full- or half-sib progenies reported a comparable performance of A-matrix and G-matrix based methods in *Pinus taeda* [18, 23], Douglas-fir [29] and *Picea mariana* [15] for growth related traits and wood properties. Moreover, ABLUP accuracies were lower for growth, form and wood quality in *Eucalyptus nitens* [24]. Experimental design factors such as number of progenies and their level of coancestry, statistical method and the traits and pedigree errors under study may account for the apparent inconsistence in the relative performance of both methods [31].

Our results indicate that for more heritable traits ABLUP and GBLUP capture similar levels of additive variance, whereas for traits with very low heritability using ABLUP, such as MFA, the markers are able to capture additional genetic variance probably in the form of historical pedigree reflected in the G matrix. Less obvious is the case for Velocity and MOE_{ind} where GBLUP seems to capture lower values of additive variance. It is possible that at intermediate values of h^2 the benefits of capturing historical consanguinity is overcome by possible confounding effects caused by markers which are identical by state (IBS) or simply due to genotyping errors. The h^2 values obtained with ABLUP and GBLUP is the result of a balance between multiple factors such as the genetic structure of the trait, the historical pedigree, and the possible model overfitting to spurious effects or genotyping errors.

Effects on GS model predictive ability (PA) of training-to-validation sets ratios and statistical methods

In conifers and *Eucalyptus* cross-validation is often performed on 9/1 training to validation sets ratio [8, 12, 15, 16, 28]. This coincide with the general conclusion from the present study, with the exception of MFA and MOE, for which the best results were obtained at ratio 8/2. It has been suggested that when the trait has large standard deviation, more training data is needed to cover the variance in order to get high predictive ability [32]. Therefore, for density, Pilodyn and Velocity, PA kept increasing with the size of the training set increased. But for other traits with smaller standard deviation, (4.44 and 2.28 for MFA and MOE), PA decreased when increasing the training set

from 80 to 90%, which may indicate that too much noise was introduced during model training.

The fact that the estimated PAs for all the solid wood properties as measured by SiliviScan are 25–30% higher than their proxies estimated from measurements of penetration depths and sound velocity at the bark may reflect the indirect nature of their proxies: the correlations calculated for the almost 6000 trees initially sampled were − 0.62 between Pilodyn and density, − 0.4 between Velocity and MFA and 0.53 between MOE_{ind} and MOE [33].

In the conifer literature it has more often been reported similar performance of different marker-based statistical models for wood properties [11, 12, 18, 28, 34]. This general conclusion agrees with our findings for all our traits with the exception of Velocity and to a less extent of MOE_{ind}. For these two traits, GBLUP and rrBLUP performed better than the other GS methods, which could be the result of a highly complex genetic structure where a large number of genes of similar and low effect are responsible for controlling of the trait. For traits affected by major genes the variable selection methods, for example BayesB or LASSO, have been reported to perform better [18], whereas for additive traits the use of nonparametric models may not yield the expected accuracy [35].

Comparison of PA and PC from methods based on pedigree and markers

Generally, pedigree-based PA estimates in conifer species have been reported to be higher or comparable to marker-based models [11, 15, 16, 19, 20, 23], but there are also some studies reporting marker-based PA estimates to be higher [13, 24, 36]. Our results for density and Pilodyn follow the general finding in forest trees, whereas for MFA, a low heritability trait, the PA estimation based on GBLUP model is substantially higher (0.16) compared to the ABLUP model (0.04). When PA is standardized with h, the predictive accuracies of the methods become more similar across traits, indicating that proportionally similar response to GS can be expected for all traits.

Use of tree age versus cambial age (ring number)

From a quick look at Fig. 2, one may get the impression that breeding based on cambial age data allows earlier selection than using tree age data. That would however be a too rushed conclusion. At tree age 3 years, after the vegetation period of 1993, only 12.5% of the trees had formed the first annual ring at breast height. Not until tree age 6 years, more than 90% of the trees had done so. But if aiming for 90% representation, one must wait several years more until more rings are formed at breast height, i.e., from 1993 to end of growth season 1996 at tree age 6. And to train models based on data from 90%

Zhou *et al. BMC Genomics*   (2020) 21:323

Page 8 of 12

of the trees for cambial age say, age 6 at breast height, samples cannot be collected until the end of growth season at tree age 11 years, or if a representation of 80% is judged as satisfactory, at tree age 10 years. This has to be considered if selection efficiencies are calculated based on cambial age data, which is common. Such results have for instance been published based on the almost 6000 trees sampled at 2011 and 2012 [37].

Correctly compared based on minimum 90% of the trees, the estimated PAs shown in Fig. 2 are similar between the age alternatives, or slightly better for use of tree age. For example, the PA for MOE using cambial age data shows a smooth increase, reaching above 0.2 at cambial age (ring number) 7, which needs data from the tree of age 12. The corresponding curve from using tree age passed above 0.2 already at age 8 years. However, curves based on tree age often show larger year-to-year variation. This is most likely an effect of the fact that the rings of same cambial age represent wood formed across a span of years with different weather. Thus, cambial age data reflect annual weather across a range of years, which does not happen when using tree age data. On the other hand, from a practical point of view, methods based on using tree age may be easier to apply in operational breeding, especially as light color results in Fig. 3b-d, indicating that high PAs can be reached without coring all the way to the pith. To number the rings for precise cambial age, you need to find the innermost ring at the pith, but that may not be necessary for good results.

Implementation of GS for solid wood into operational breeding

The results indicate that GS can result in similar early selection efficiency or even higher than traditional pedigree-based breeding and offers further possibilities. Previously, in loblolly pine it was reported that models developed for diameter at breast height (DBH) and height with data collected on 1 to 4-year old trees had limited accuracy in predicting phenotypes at age 6-year old [21]. In British Columbia Interior spruce, the predictive accuracy for tree height of models trained at ages 3 to 40 years, at certain intervals, and validated at 40 years revealed less opportunities for early model training, since the plateau was not reached until 30 years [28].

In our study, the highest PA values (on the diagonals in Fig. 3b-d) were obtained for the subsets of fast-growing trees which had reached breast height already at tree age 3 and 4 years, 12 and 51% of the total number of trees, representing a limited number of the OP families included in the analysis. Trees in this subgroup are affected by high intensity of selection for alleles accelerating growth within each OP family. Also, on cross-validation the prediction abilities for this group were calculated based on the trees within the same group. In this elite group different factors could account for a higher PA value, such as lower phenotypic variance, decreased number of alleles of minor effect could also facilitate identification of major effects and/or higher consanguinity between those families which may share alleles for growth. These models are shown for completeness, but as they cannot be used for operational breeding they are not further discussed.

Models for genetic selection are useful in different steps of a breeding program. One type of prediction models, here illustrated with Table 1, can be trained from existing trials, preferably based on trees of as old age as available. Since the aim of breeding is to predict tree qualities at age of harvesting when the major part of the stem will be dominated by mature wood. Training the models in older trees for wood properties also allows considering other properties which cannot be easily observed from trees of very young age, such as stem straightness and health. For wood density, the results indicate that models can be built without coring very deep into the stem. It may be expected that this is valid also for instance for tracheid dimensions which in combination determines the wood density [30].

As illustrated in this work, two aspects of incorporating wood properties into operational GS breeding programs can be addressed with the same set of data. Firstly, as mentioned above, models for cost-effective selection based on genomic information from existing trees. In that case, models from data at old ages would normally be preferred, for example for wood density some model at bottom line of Fig. 3b. Secondly, models providing guidance on at what age it is reasonable to approach young trees for training of GS models for specific traits: a) trees in existing juvenile trials, or b) trees of new generations with different pools of genetics. As an example, the same Fig. 3b for wood density suggests GS model training at tree ages 10 to 12 on the third to fifth outermost rings to reduce costs and the negative impact on the tree.

Conclusions

1) In comparison with phenotypic selection, Genomic selection methods showed similar to higher prediction abilities (PAs) for both increment core- and standing tree-based phenotyping methods. This indicates that the standing tree-based measurements may be a cost-effective alternative method for GS, but higher PAs were obtained based on increment core-based wood analyses.

2) Different genomic prediction statistical methods provided similar PA. At least 80% data should be

included in the training set in order to reach the highest levels of PA

3) This study represents the first published investigation of the efficiency of GS with prediction models trained on data acquired from sampling/coring trees at different ages, combined with sampling/coring to different depths, to optimize the operational breeding for the combination of length of breeding cycle, cost and impact on the trees. The results indicate that similar efficiency can be obtained at tree age 10–12 with 3–5 outermost rings.

Methods

Plant material

The study was conducted on two OP progeny trials: S21F9021146 (F1146) (Höreda, Eksjö, Sweden) and S21F9021147 (F1147) (Erikstorp, Tollarp, Sweden). Both trials were established in 1990 with a spacing 1.4 m × 1.4 m. Originally, the experiments contained more than 18 progenies from 524 families at each of site, but after thinning activities in Höreda and Erikstorp in 2010 and 2008, respectively, about 12 progenies per family were left. In 2011 and 2012, six trees per site (524 * 12 ~ 6000 trees) were phenotyped [37]. Standing tree-based measurements with Pilodyn and Hitman were performed on the same trees in 2011 and 2013, respectively, after which further thinning was performed. For this study, in 2018, we generated genomic (SNP) data from 484 remaining progeny trees after thinning which belonged to 62 of the OP families (out of the original 524 families) and on average eight progenies per family. This genotypic data was combined with available phenotypic data for the same trees that were used.

Phenotypic data

The phenotypic data was previously described in Zhou et al., 2019 [38]. Increment cores of 12 mm diameter from pith to bark were collected from the progenies in 2011 and in 2012. These samples were analyzed for pith to bark variations in many woods and fiber traits with a SilviScan [39] instrument at Innventia (now RISE), Stockholm, Sweden. This data is referred as increment core-based measurements through the text. The annual rings of all samples were identified, as well as their parts of earlywood, transition wood and latewood, averages were calculated for all rings, as well as their parts and dated with year of wood formation [30].

The aim of breeding is not for properties of individual rings, but properties of the stem at harvesting target age. Therefore, this study focused on predictions of averages for stem cross-sections, and we chose tree age 19 years as the reference age, with models trained on trait averages for all rings formed up to different younger ages.

Three types of averages were calculated and predictions compared for density, MFA and MOE: 1) area-weighted averages, relating to the cross-section of the stem, 2) width-weighted, relating to a radius or an increment core, and 3) arithmetic averages, where all ring averages are weighted with same weight. For the calculation of area-weighted average we assumed that each growth ring is a circular around the pith, calculated the area of each annual ring from its inner and outer radii, and when calculating the average at a certain age, the trait average for each ring was weighted with the ring's proportion of the total cross-sectional area at that age. Similarly, for the calculation of the width-weighted average, the trait average for each ring was weighted with the ring's proportion of the total radius from pith to bark at that age. Similar results were obtained with the three average methods. For this reason, only the estimates based on the area-weighted method (the most relevant for breeding) are shown. Tree age 19 years was used as the reference age. Thus, all the selection methods investigated for density, MFA and MOE, phenotypic and genetic, were compared based on how well they predicted the cross-sectional averages of the trees at this age, with their last ring formed during the vegetation period of 2009.

In addition, estimates of the three solid wood traits were calculated based on data from Pilodyn and Hitman instruments, measured on the standing trees without removing the bark at age 22 and age 24 years, respectively. Pilodyn measures the penetration depth with a needle pressed into the stem, which is inversely correlated with wood density. Hitman measures the velocity of sound in the stem, which correlates with microfibril angle, MFA [40, 41]. MOE is related to wood density and velocity of sound [42–44] and can therefore be estimated by combining the Pilodyn and Velocity data, which estimates we here name MOE_{ind} (for standing-tree based). Further details on how this was performed in our study are given in Chen et al. 2015 [33]. The references show that these standing-tree-based measurements provide useful information and are very time and cost-efficient. However, they do not allow calculation of properties of the tree at younger ages. Therefore, we were not able to investigate from what early ages such data can be uses within genomic selection.

Genotypic data

Genomic DNA was extracted from buds or needles when buds were not available. Qiagen Plant DNA extraction protocol was utilized for DNA extraction and purification and DNA quantification performed using the Qubit® ds DNA Broad Range (BR) Assay Kit (Oregon, USA). Genotyping was conducted at Rapid

Genomics, USA, using exom capture methodology same as the method used in Baison et al. 2019 [45]. Sequence capture was performed using the 40,018 diploid probes previously designed and evaluated for *P. abies* [46] and samples were sequenced to an average depth of 15x using an Illumina HiSeq 2500 (San Diego, USA) [45]. Variant calling was performed using the Genome Analysis Toolkit (GATK) HaplotypeCaller v3.6 [47] in Genome Variant Call Format (gVCF) output format. After that, the following steps were performed for filtering: 1) removing indels; 2) keeping only biallelic loci; 3) removing variant call rate ("missingness") < 90%; 4) removing minor allele frequency (MAF) < 0.01. Beagle v4.0 [48] was used for missing data imputation. After these steps, 130,269 SNPs were used for downstream analysis.

Population structure
As a first step, we conducted a principal component analysis to determine the presence of structure in our population. The spectral decomposition of the marker matrix revealed that only about 2% of the variation was captured by the first eigenvector, indicating low population structure. Additionally, in previous study, low genotype by environment (GxE) interaction was detected for wood quality traits on these two trials [37]. Therefore, population structure was not considered in the design of cross-validation sets (see Modelling and cross-validation chapter for further details on the cross-validation sets design).

Narrow-sense heritability (h²) estimation
For each trait, an individual tree model was fitted in order to estimate additive variance and breeding values:

$$y = X\beta + Zu + Wb + e. \tag{1}$$

where y is a vector of measured data of a single trait, β is a vector of fixed effects including a grand mean, provenance and site effect, b is a vector of post-block effects and u is a vector of random additive (family) effects which follow a normal distribution $u \sim N(0,A\sigma^2_u)$ and e is the error term with normal distribution $N(0,I\sigma^2_e)$. X, Z and W are incidence matrices, A is the additive genetic relationship matrix and I is the identity matrix. σ^2_u equals to σ_a^2 (pedigree-based additive variance) when random effect in eq. 1 is pedigree-based in which case $u \sim N(0,A\sigma^2_u)$, and σ^2_u equals to σ_g^2 (marker-based additive variance) when random effect in eq. 1 is marker-based in which case $u \sim N(0,G\sigma^2_u)$. The G matrix is calculated as $G = \frac{(M-P)(M-P)T}{2\sum_{i=1}^{q} pi(1-pi)}$, where M is the matrix of samples with SNPs encoded as 0, 1, 2 (i.e., the number of minor alleles), P is the matrix of allele frequencies with the ith column given by 2(pi – 0.5), where pi is the observed allele frequency of all genotyped samples.

Pedigree-based individual narrow-sense heritability (h_a^2) and marker-based individual narrow-sense heritability (h_g^2) were calculated as.

$$h_a^2 = \frac{\sigma_a^2}{\sigma_{pa}^2}, \quad h_g^2 = \frac{\sigma_g^2}{\sigma_{pg}^2}$$

respectively, σ^2_{pa} and σ^2_{pg} are phenotypic variances for pedigree-based and marker-based models, respectively.

Selection of the optimal training and validation sets ratio
Cross-validation was conducted after dividing randomly the whole dataset into a training and a validation set. To find the most suitable ratio between the two, we divided the data into sets with five different ratios between the training and the validation sets: 50, 60, 70, 80 and 90%. 100 replicate iterations were carried out for each tested ratio and trait.

Statistical method for model development
In the same context we aimed to find optimal methods. Several statistical methods were compared: pedigree-based best linear unbiased predictions (ABLUP), and four GS methods: genomic best linear unbiased predictions (GBLUP) [49], random regression-best linear unbiased predictions (rrBLUP) [4, 50], BayesB [4], and reproducing kernel Hilbert space (RKHS).

rrBLUP used a shrinkage parameter lamda in a mixed model and assumes that all markers have a common variance. In BayesB the assumption of common variance across marker effects was relaxed by adding more flexibility in the model. RKHS does not assume linearity so it could potentially capture nonadditive relationships [51]. R package *rrBLUP* [52] was used for GBLUP and rrBLUP, package *BGLR* [53] was used for BayesB and RKHS. The pedigree-based relationship matrix was obtained with the R package *pedigree* [54].

PA and accuracy estimation
The adjusted phenotypes $y' = y-X\beta$ were used as model response in the genomic prediction models. Model quality was evaluated by predictive ability (PA), which is the mean of the correlation between the adjusted phenotype and the model predicted phenotypes, r(y',yhat) from 100 times CV. Prediction accuracy (PC) was defined as PA/ $\sqrt{(h^2)}$ [15, 55]. In order to investigate whether GS model training can be conducted at earlier age, PA at each tree calendar age and cambial age were estimated. In this case, cross validation was conducted only using area-weighted values at each age, then the trait values at each age were estimated. PA at a specific age was calculated as the correlation between estimated trait values at that age and area-weighted values from pith to the last ring

Zhou *et al. BMC Genomics*   (2020) 21:323

Page 11 of 12

(for cambial age) and last year (for calendar age), respectively.

Genomic selection for well-performing trees with the use of marker information (G matrix) requires access to previously trained GS models. Thus, model training is a necessary part of GS integration into operational breeding. Model training can be conducted in already existing plantations with trees of relatively high ages, as illustrated in this work. It is, however, expected and desired that such model training can be conducted with high PAs also for younger trees. This would be especially useful if maturity (flower production) can be accelerated, to shorten the total breeding cycle.

Operationally, it is also important to develop protocols to assess wood quality in resources at minimum cost and time, and with minimal impact on the trees. Therefore, on coring, it is not only important to know the minimum age at which useful information can be obtained, but also from how many rings from the bark towards the pith information is required to train models with high predictive ability. To address these two practical questions for operational breeding, we trained prediction models based on data from different sets of rings, in order to mimic and compare PAs obtained when coring at different ages of the trees to different depths into the stem, or more precisely, using data from different numbers of rings, starting next to the bark. All the models were judged on, compared by their ability to predict the cross-sectional average of the trait at age 19 years across all trees in the validation set.

Abbreviations
ABLUP: Pedigree-based best linear unbiased prediction; BR: Broad Range; DBH: Diameter at breast height; GATK: Genome Analysis Toolkit; GBLUP: Genomic best linear unbiased predictions; GS: Genomic selection; GxE: Genotype by environment interaction; gVCF: Genome Variant Call Format; IBS: Identical by state; LD: Linkage disequilibrium; MAF: Minor allele frequency; MFA: Microfibril angle; MOE: Modulus of elasticity; PA: Predictive ability; PC: Prediction accuracy; QTL: Quantitative trait loci; OP: Open-pollinated; rrBLUP: Random regression-best linear unbiased predictions; RKHS: Reproducing kernel Hilbert space

Authors' contributions
LZ analysed data and drafted the manuscript. ZC designed sampling strategy, coordinated field sampling and edited the manuscript. BK participated in the selection of the breeding populations, providing access to field experiments and edited the manuscript. LO, TG conducted the SilviScan measurements and performed the evaluations prior to the genetic analyses. HW conceived and designed the study and edited manuscript. SOL and RRG provided ideas and revised manuscript. All authors read and approved the final manuscript.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
The plant materials analysed for this study comes from common garden experiments that were established and maintained by the Forestry Research Institute of Sweden (Skogforsk) for breeding selections and research purposes. Three tree breeders in Sweden were co-authors in this paper. They agreed to access the materials.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Forest Genetics and Plant physiology, Umeå Plant Science Centre, Swedish University of Agricultural Sciences, SE-901 83 Umeå, Sweden. [2]RISE Bioeconomy, Box 5604, SE-114 86 Stockholm, Sweden. [3]Skogforsk, Ekebo 2250, SE-268 90 Svalöv, Sweden. [4]Beijing Advanced Innovation Centre for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing, China. [5]CSIRO National Collection Research Australia, Black Mountain Laboratory, ACT, Canberra 2601, Australia. [6]IIC, Rosenlundsgatan 48B, SE-118 63, Stockholm, Sweden.

References
1. Hannrup B, et al. Genetic parameters of growth and wood quality traits in Picea abies. Scand J For Res. 2004;19(1):14–29.
2. Erickson U. Skogforsk, Strategi för framtida skogsträdsförädling och framställning av förädlat skogsodlingsmaterial i Sverige; 1995.
3. Karlsson B, Rosvall O. Progeny testing and breeding strategies. Proceedings of the Nordic group for tree breeding. Edinburgh: Forestry commission; 1993.
4. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157(4):1819–29.
5. Bouvet J-M, et al. Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping: prediction accuracy implications. Heredity. 2016;116(2):1365–2540.
6. El-Dien OG, et al. Implementation of the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from nonadditive genetic effects. G3: genes. Genomes, Genetics. 2016;6(3):743–53.
7. El-Kassaby YA, et al. Breeding without breeding: is a complete pedigree necessary for efficient breeding? PLoS One. 2011;6(10):e25737.
8. Munoz PR, et al. Genomic relationship matrix for correcting pedigree errors in breeding populations: impact on genetic parameters and genomic selection accuracy. Crop Sci. 2014;54(3):1115–23.
9. Tan B, et al. Genomic relationships reveal significant dominance effects for growth in hybrid Eucalyptus. Plant Sci. 2018;267:84–93.
10. Grattapaglia D, Resende MDV. Genomic selection in forest tree breeding. Tree Genet Genomes. 2011;7(2):241–55.
11. Chen ZQ, et al. Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. BMC Genomics. 2018;19(1):946.
12. Isik F, et al. Genomic selection in maritime pine. Plant Sci. 2016;242:108–19.
13. Kainer D, et al. Accuracy of Genomic Prediction for Foliar Terpene Traits in Eucalyptus polybractea. G3: genes, Genomes, Genetics. 2018;8(8):2573–83.
14. Zapata-Valenzuela J, et al. SNP markers trace familial linkages in a cloned population of Pinus taeda—prospects for genomic selection. Tree Genet Genomes. 2012;8(6):1307–18.
15. Lenz PRN, et al. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (Picea mariana). BMC Genomics. 2017;18(1):335.

16. Müller D, Schopp P, Melchinger AE. Persistency of Prediction Accuracy and Genetic Gain in Synthetic Populations Under Recurrent Genomic Selection. G3: Genes Genomes Genetics. 2017;**7**(3):801–11.

17. Tan B, et al. Evaluating the accuracy of genomic prediction of growth and wood traits in two Eucalyptus species and their F 1 hybrids. BMC Plant Biol. 2017;17(1):110.

18. Resende MFR Jr, et al. Accuracy of genomic selection methods in a standard data set of loblolly pine (Pinus taeda L.). Genetics. 2012c;190(4): 1503–10.

19. Beaulieu J, et al. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. Heredity. 2014a; 113(4):343–52.

20. El-Dien OG, et al. Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. BMC Genomics. 2015;16(1):370%@ 1471–2164.

21. Resende MFR Jr, et al. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytol. 2012b;193(2):617–24.

22. Resende MDV, et al. Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol. 2012;194(1):116–28.

23. Zapata-Valenzuela J, et al. Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. G3: Genes Genomes, Genetics. 2013;3(5):909–16.

24. Suontama M, et al. Efficiency of genomic prediction across two Eucalyptus nitens seed orchards with different selection histories. Heredity. 2019;122(3):370.

25. Beaulieu J, et al. Genomic selection accuracies within and between environments and small breeding groups in white spruce. BMC Genomics. 2014b;15(1):1048.

26. Daetwyler HD, et al. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics. 2013;193(2):347–65.

27. de Almeida Filho JE, et al. The contribution of dominance to phenotype prediction in a pine breeding and simulated population. Heredity. 2016; 117(1):33–41.

28. Ratcliffe B, et al. A comparison of genomic selection models across time in interior spruce (Picea engelmannii x glauca) using unordered SNP imputation methods. Heredity. 2015;115(6):547–55.

29. Thistlethwaite FR, et al. Genomic selection of juvenile height across a single-generational gap in Douglas-fir. Heredity. 2019;122(6):848–63.

30. Lundqvist S-O, et al. Age and weather effects on between and within ring variations of number, width and coarseness of tracheids and radial growth of young Norway spruce. Eur J For Res. 2018;137(5):719–43.

31. Vela-Avitúa S, et al. Accuracy of genomic selection for a sib-evaluated trait using identity-by-state and identity-by-descent relationships. Genet Sel Evol. 2015;47(1):9.

32. Isidro J, et al. Training set optimization under population structure in genomic selection. TAG. 2015;128(1):145–58.

33. Chen Z-Q, et al. Estimating solid wood properties using Pilodyn and acoustic velocity on standing trees of Norway spruce. Ann For Sci. 2015; 72(4):499–508.

34. Thistlethwaite FR, et al. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. BMC Genomics. 2017;18(1):930.

35. Desta ZA, Ortiz R. Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci. 2014;19(9):592–601.

36. El-Dien OG, et al. Multienvironment genomic variance decomposition analysis of open-pollinated interior spruce (Picea glauca x engelmannii). Mol Breed. 2018;38(3):26.

37. Chen Z-Q, et al. Inheritance of growth and solid wood quality traits in a large Norway spruce population tested at two locations in southern Sweden. Tree Genet Genomes. 2014;10(5):1291–303.

38. Zhou L, et al. Genetic analysis of wood quality traits in Norway spruce open-pollinated progenies and their parent plus trees at clonal archives and the evaluation of phenotypic selection of plus trees. Can J For Res. 2019;49(7):810–8.

39. Evans R. Rapid Measurement of the Transverse Dimensions of Tracheids in Radial Wood Sections from Pinus radiata. Holzforschung: International Journal of the Biology, Chemistry, Physics and Technology of Wood; 1994. p. 168.

40. Downes GM, et al. Relationship between wood density, microfibril angle and stiffness in thinned and fertilized Pinus radiata. IAWA J. 2002;23(3):253–65.

41. Lenz P, et al. Genetic improvement of white spruce mechanical wood traits—early screening by means of acoustic velocity. Forests. 2013;4(3):575–94.

42. Haines DW, Leban J-M. Evaluation of the MOE of Norway spruce by the resonance flexure method. For Prod J. 1997;47(10):91.

43. Knowles RL, et al. Evaluation of non-destructive methods for assessing stiffness of Douglas fir trees. N Z J For Sci. 2004;34(1):87–101.

44. Lindström H, Harris P, Nakada R. Methods for measuring stiffness of young trees. Holz als Roh-und Werkstoff. 2002;60(3):165–74.

45. Baison J, et al. Genome-wide association study identified novel candidate loci affecting wood formation in Norway spruce. Plant J. 2019;100(1):83–100.

46. Vidalis A, et al. Design and evaluation of a large sequence-capture probe set and associated SNPs for diploid and haploid samples of Norway spruce (Picea abies). bioRxiv. 2018:291716..

47. McKenna A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9): 1297–303.

48. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007;81(5):1084–97.

49. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91(11):4414–23.

50. Whittaker JC, Thompson R, Denham MC. Marker-assisted selection using ridge regression. Genet Res. 2000;75(2):249–52.

51. Heslot N, et al. Genomic selection in plant breeding: a comparison of models. Crop Sci. 2012;52:146–60.

52. Endelman JB. Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome. 2011;4:250–5.

53. Pérez P. And G. de los Campos, *Genome-wide regression and prediction with the BGLR statistical package*. Genetics. 2014;198(2):483–95.

54. Coster A. pedigree: Pedigree functions. R package version; 2013. p. 1.

55. Dekkers JCM. Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet. 2007;124(6): 331–41.

Publisher's Note

VI

Genetic control of tracheid properties in Norway spruce wood

Baison J[1], Linghua Zhou[1], Nils Forsberg[1], Tommy Mörling[1], Thomas Grahn[2], Lars Olsson[2], Bo Karlsson[4], Harry X Wu[1], Ewa J. Mellerowicz[1], Sven-Olof Lundqvist[2,3], and María Rosario García-Gil[1]*

Corresponding author: m.rosario.garcia@slu.se

1. Department of Forest Genetics and Plant Physiology, Umeå Plant Science Centre, Swedish University of Agricultural Science, Umeå, Sweden

2. RISE Bioeconomy, Box 5604, SE-114 86 Stockholm, Sweden

3. IIC, Rosenlundsgatan 48B, SE-11863 Stockholm, Sweden

4. Skogforsk, Ekebo 2250, SE-268 90, Svalov, Sweden

Abstract

Through the use of genome-wide association studies (GWAS) mapping it is possible to establish the genetic basis of phenotypic trait variation. Our GWAS study presents the first such effort in Norway spruce (*Picea abies* (L). Karst.) for the traits related to wood tracheid characteristics. The study employed an exome capture genotyping approach that generated 178 101 Single Nucleotide Polymorphisms (SNPs) from 40 018 probes within a population of 517 Norway spruce mother trees. We applied a least absolute shrinkage and selection operator (LASSO) based association mapping method using a functional multi-locus mapping approach, with a stability selection probability method as the hypothesis testing approach to determine significant Quantitative Trait Loci (QTLs). The analysis has provided 30 loci and 26 candidate genes, the majority of which show specific expression in wood-forming tissues or high ubiquitous expression, potentially controlling tracheids dimensions, their cell wall thickness and microfibril angle. Among the most promising candidates based on our results and prior information for other species are: *Picea abies BIG GRAIN 2* (*PabBG2)* with a predicted function in auxin transport and sensitivity, and *MA_373300g0010* encoding a protein similar to wall-associated receptor kinases, which were both associated with cell wall thickness. The results demonstrate feasibility of GWAS to identify novel candidate genes controlling industrially-relevant tracheid traits in Norway spruce

Introduction

Norway spruce is considered to be one of the most important multipurpose species. Its wood provides various solid wood products as well as pulp and paper products. It is considered one of the best raw-materials for the production of mechanical pulp for many types of paper grades[1]. The properties of the tracheids have large influences on the quality of the final products, and also on process economy and sustainability, for solid wood as well as fibre-based products[2]. Tracheid morphology and cell wall structure influence the flexibility of wood and fibres, interactions among fibres, as well as the mechanical, physical and optical properties of the end-products[3]. Consequently, identifying the genetic background of different tracheid traits as a basis for breeding may bring benefits for both industry and society. Several papers have reported the phenotypic correlations, between tracheid cross-sectional

dimensions and wood traits such as density in conifers[4,5,6]. A study of Norway spruce felled in the winter of 1989/1990 in central Sweden, found tracheid length dependent on the logarithm of cambial age and growth ring width, with density dependent on latewood percentage. Similar models for the influence of cambial age and ring width have been presented for tracheid length, width and wall thickness models of Norway spruce, Sitka spruce, Scots pine and loblolly pine[7]. Such results have indicated that changes in growth conditions over time acting mainly through crown development, will have an influence on wood structure development in Norway spruce[6]. However, these reports paid very little regard to the underlaying genetic factors influencing these phenotypes. Therefore, the dissection of the genetics impacting these relationships and the variations observed in tracheid properties will be of great value to any tree breeding program.

Various long-term breeding programmes for the species are already being pursued with the goal to identify genotypes with high productivity and wood quality[8]. Wood density and microfibril angle (MFA) are key indicators of wood quality as they influence strength and dimensional stability of solid wood[9]. However, combining productivity with wood quality is problematic due to negative genetic correlations between these traits[10]. One of the tools helping to understand these genetically complex variations in forest trees is the integration of extensive genetic and phenotypic data in order to discern the genetics underlying these traits[11,12,13]. Hence, knowing the genetic control of these variations, may lead to optimal breeding strategies for the improvement of both growth and wood quality traits.

With genomic resources now available, a large array of molecular markers has been available for the studying and understanding of complex traits. The majoritiy of these traits are known to be predominantly polygenic in nature, and affected by environmental effects[14], hence the need to utilize techniques that target the whole genome[15]. The availability of an array of genomic resources has led to the reliable identification of Quantitative Trait Loci (QTLs), which in conifers are traditionally detected using suitable segregating populations such as, full- or half-sib progenies. More recently, GWAS, also known as Linkage Disequilibrium (LD) mapping, has been applied as an alternative approach of QTL detection from traditional pedigree-based mapping studies. GWAS accounts for historical recombination events in the natural population as compared to those observed in a pedigree

based QTL mapping[16]. When confounding factors are taken into consideration, LD mapping provides greater resolution than pedigree studies, since it utilizes markers in strong LD with putative causative genomic regions[17].

Many coniferous species are characterized by an outcrossing mating system and large population sizes which lead to a rapid LD decay within the genomes and low inbreeding coefficient[16]. However, rapid and heterogenous decay in conifer LD[18] can be a source of concern as proximal markers can be completely unlinked and therefore offer no predictive power to the quantitative trait that may be residing physically close[19]. Together LD heterogeneity, population structure[20], epistasis and Genotype x Environment interactions (GxE)[21] are factors that if not carefully controlled can negatively impact QTL identification. The utilization of LD mapping in the dissection of genetic backgrounds underpinning complex traits has been shown in several systems, for example, complex solid wood properties in Norway spruce[22], white spruce[23] and *Eucalyptus*[19], and detecting genes underlying ecological adaptations in *Populus*[24]. The dissection of these complex traits can benefit from the application of mathematical functions that account for the year-to-year variation across annual growth rings. The development of mathematical methods for the analysis of dynamic data has made it possible to develop functional mapping approaches[25,26] that firstly model the phenotypes using curve-fitting methods and then utilize the parameters describing the curve (latent traits) for independent association analysis[27,28].

GWAS can also increase our knowledge on molecular processes controlling tracheid traits. Presently the majority of breeding programs have focused on the easy to measure phenotypic traits such as volume, straightness, disease resistance and spiral grain. Due to cost and time of measurement of traits related to tracheid dimensions most programs have not been able to select and advance such traits using marker assisted breeding[29].Therefore, this study is novel in that it is, to our knowledge, one of the first to tackle the issue of dissecting the genetic background to tracheid properties in a conifer species. With the exception of a single study conducted in *Arabidopsis thaliana*, as a model system, for traits controlling fibre length[30], the majority of the studies related to tree fibre related traits have focused on mostly microfibril angle genetics[31,32,33,34]. Hence our study seeks to form the bases upon which, the dissection of the genetic backgrounds to more complex and expensive traits, such as,

tracheid dimensions can be investigated. Such traits are to a large extent determined by the genes acting during wood development[35,36]. Tracheid traits can also be regulated non-cell-autonomously by processes that take place in other organs and tissues. For example, the activity of the shoot apical meristem determines the availability of auxin in the cambium and developing wood[37,38], whereas the photosynthetic activity in the needles influences the availability of sucrose for wood biosynthesis[39]. Therefore, combining the knowledge of candidate genes with their expression analysis will give more insights to the biological processes shaping tracheids.

The major goal of this study was to identify causative allelic effects of genomic regions contributing to wood tracheid traits using LD mapping on exome sequence capture data. Due to the large size of the Norway spruce genome (20 Gb) and its highly repetitive nature, it presents a challenge to use whole genome re-sequencing approaches for the development of molecular markers. Approaches aimed at reducing these genome complexities, especially by either eliminating or drastically reducing the repetitive sequences have been developed[40]. These approaches are referred to as reduced representation approaches as there are used as proxies for whole genomic sequencing. In this study, we have used exome capture, aiming at maximizing the capture of exonic regions of the genome only, thereby increasing the coverage and depth of genic sequence in our variant detection study. The analysis provided 26 mostly novel candidate genes for regulation of various tracheid traits, which, along with their expression patterns, give new insights to the tracheid traits determination, and offer key markers for early genetic selection in Norway spruce breeding.

Materials and Methods

Association mapping population

The association mapping population, phenotypic data and statistical analysis are described in Chen *et al.,* (2014)[41] and Hayatgheibi *et al.,* (2018)[42]. Briefly, the mapping population for the association mapping population constisted of two progeny trails established 1990 in Southern Sweden: (S21F9021146 aka F1146 (trial1) and S21F9021147 aka F1147 (trial2)), composed of 1373 and 1375 half-sib families. A randomized incomplete block design with single-tree plots was employed for both trials. From the trials, 517 families in 112 provenances were selected for use in the investigation of

wood tracheid properties. Wood increment cores with diameter of 12 mm were collected at breast

height (1.3 m) from six trees from each of the selected families of each trial. A total of 5618 trees were

sampled: 2973 trees from trial F1146 and 2645 from F1147.

Phenotypic data generation

The radial variations of growth, wood and tracheid attributes from pith to bark were analysed using

the SilviScan instrument[43] at Innventia, now RISE Bioeconomy, Stockholm, Sweden. SilviScan is an

instrument for efficient measurement of radial variations in a multitude of properties from the same

sample with high spatial resolution. High precision sample strips from pith to bark were produced

from the increment cores and automatically scanned for radial variations in cross-sectional tracheid

widths with a video microscope combined with image analysis, in wood density with X-ray

transmission and in structural orientations with X-ray diffraction. From these data, information on

radial variations of further traits were derived, such as wall thickness, coarseness and MFA of

tracheids, and stiffness of wood (MOE). The locations of the annual rings were identified, as well as

of their compartments of earlywood (EW), transitionwood (TW) and latewood (LW), using the "20-80

density" definition[44], established for use in different types of studies[45,46,47]. Averages for all rings and

their compartments were calculated for the traits and organised to be ready for use in continued

genetic evaluations, such as the work on solid wood traits[48], on tracheid traits[49] and for wood traits[22],

genomic selection[50] and influences of age and weather[51]. The traits addressed in the current work are

listed in Table 1.

For MFA, central peak regression mathematical functions were fitted to describe the MFA

variation from juvenile towards mature wood, using procedures presented by Hayatgheibi et al.,

(2018)[52], including also pre-processing of the data for removal of outliers. A threshold value of MFA

20° for the fitted curves was chosen to define an age up to to which an inner core of wood with

inferior timber properties occurred, here named the transition age MFA_{TA}[42]. From anatomical

perspective, a threshold of 20° is on the high side, emphasizing a core of pronounced juvenility. We

have decided to stay with this threshold level, because for the young trees investigated, the fitted

curves for quite a few trees would not pass a low treshold, and they would have had to be discarded

6

from the analysis. Thus, it works better for ranking. The averages of MFA for wood inside and outside this limit were calculated, MFA$_{CORE}$ and MFA$_{OUTER}$. This provided three latent traits for MFA.

Table 1. List of the traits, their abbreviations and measurement unit.

Trait	Abbreviation	Unit
Radial tracheid width (TWr)		
Ring	TWr$_{Ring}$	μm
Earlywood	TWr$_{EW}$	μm
Transitionwood	TWr$_{TW}$	μm
Latewood	TWr$_{LW}$	μm
Tangential tracheid width (TWt)		
Ring	TWt$_{Ring}$	μm
Earlywood	TWt$_{EW}$	μm
Transitionwood	TWt$_{TW}$	μm
Latewood	TWt$_{LW}$	μm
Wall Thickness (WT)		
Ring	WT$_{Ring}$	μm
Earlywood	WT$_{EW}$	μm
Transitionwood	WT$_{TW}$	μm
Latewood	WT$_{LW}$	μm
Coarseness (C)		
Ring	C$_{Ring}$	mg/m
Earlywood	C$_{EW}$	mg/m
Transitionwood	C$_{TW}$	mg/m
Latewood	C$_{LW}$	mg/m
Microfibril angle (MFA)		

Ring	MFA$_{Ring}$	Degrees
Corewood	MFA$_{CORE}$	Degrees
Outerwood	MFA$_{OUTER}$	Degrees
Transition age (cambial)	MFA$_{TA}$	Year

Exome Capture Analysis

DNA extraction, variant detection and annotation and population structure on the genomic data utilized in this study was previously described[22]. Total genomic DNA from 517 half-sib individuals was extracted using the Qiagen Plant DNA extraction kit (Qiagen, Hilden, Germany). DNA was extracted from buds, when present, or from young needles, when buds were absent. DNA quantification was performed using the Qubit® ds DNA Broad Range (BR) Assay Kit (Oregon, USA). DNA from randomly selected individuals was then electrophoresed on a 2% agarose gel. Probe design and evaluation is described in Vidalis *et al.,* (2017)[53]. In breif, the exome capture method was implemented by the probe design that was based on a combination of sequenced genomic DNA, predicted gene annotations and *de novo* transcript assemblies. Exome capture was based upon the use of targeted oligonucleotides that bind to complementary genomic sequences. Sequencing was performed at Rapid Genomics, USA, using the Illumina sequencing platform. Sequence capture with average depth of 15x coverage was performed using the 40 018 diploid probes previously designed and evaluated for Norway spruce. Illumina sequencing compatible libraries were amplified with 14 cycles of PCR and the probes were then hybridized to a pool comprising 500 ng of 8 equimolarly combined libraries following Agilent's SureSelect Target Enrichment System (Agilent Technologies). These enriched libraries were then sequenced to an average depth of 15x using an Illumina HiSeq 2500 (San Diego, USA) on the 2 x 100 bp sequencing mode at Rapid Genomics, USA.

Raw reads were mapped against the *P. abies* reference genome v1.0 using BWA-MEM[54]. SAMTools v.1.2[55] and Picard (http://broadinstitute.github.io/picard) were used for sorting and marking of PCR duplicates. Variant calling was performed using GATK HaplotypeCaller v.3.6 as per the best practices protocol[56] in gVCF output format (see

for more information about GATK best practices). Samples were then merged into batches of ~200 before all 517 samples were jointly called.

GATK based Variant Quality Score Recalibration (VQSR) method was performed in order to avoid the use of hard filtering for exome/sequence capture data. For the VQSR analysis, two datasets were created: a training file and an input file. The training dataset was derived from a Norway spruce genetic mapping population with known segregating loci. The training dataset was designated as true SNPs and assigned a prior value of 12.0. The input file was derived from the raw sequence data using the above mentioned GATK's best practices with the following parameters: extended probe coordinates by +100 excluding INDELS, excluding the LowQual sites, and keeping only bi-allelic sites. The annotation parameters QualByDepth (QD), MappingQuality (MQ) and BaseQRankSum, with tranches 100, 99.9, 99.0 and 90.0, were then applied to the two files for the determination of the good versus bad variant annotation profiles. After obtaining a VQSR for all raw data variant sites, the recalibration was applied to filter the raw variants. The SNP trimming and cleaning involved the removal of any SNP with MAF and "missingness" of < 0.05 and >20%, respectively. These parameters were filtered out using VCFTools[57]. The resultant SNPs were annotated using default parameters for snpEff 4[58]. Ensembl general feature format (GTF, gene sets) information was utilized to build *P. abies* 1.0 snpEff database.

GWAS LASSO

Latent traits expressing how the traits developed with age were calculated in two steps. First, a breeding value approach was applied to refine data from influences not directly related to the genes, such as site and block effects. For this purpose, breeding values were estimated (EBV) for each annual ring separately (cambial age), reducing site and block effects, but also the time trajectories, which were reconstructed as a final step by adding back the averages at each age. The variance and covariance components were estimated using ASREML 4.0 as described in Chen *et al*., (2014)[10]. The EBVs at each cambial age were estimated using univariate, bivariate or multivariate mixed linear models in order to select the optimal model for each trait, based on a compromise of model fit and complexity. Akaike Information Criteria (AIC) was used to determine the fitness of different models.

This resulted in use of a univariate linear mixed model for joint-site analysis as the bases for the analyses of all traits:

$$Y_{ijkl} = u + S_i + B_{j(i)} + F_k + SF_{ik} + e_{ijkl} \qquad [1]$$

where Y_{ijkl} is the observation on the lth tree from the kth family in jth block within the ith site, u is the general mean, S_i and $B_{j(i)}$ are the fixed effects of the ith site and the jth block within the ith site, respectively, F_k and SF_{ik} are the random effects of the kth family and the random interactive effect of the ith site and kth family, respectively, e_{ijkl} is the random residual effect.

For the tracheid dimension and coarseness traits, linear splines with multiple knots were fitted to the EBV refined time trajectories against cambial age (annual ring number) (Fig 1), generally defined as follows:

$$y(t) = \beta_0 + \beta_1 t + \beta_2 (t - K_1)_+ + \beta_3 (t - K_2)_+ + \ldots + \beta_{1+m} (t - K_m)_+, \qquad [2]$$

This is a continuous curve starting at the intercept β_0, with linear segments between the knots at $t=K_i$ ($i=1,\ldots,m$; $K_1<K_2\ldots<K_m$), segments with slopes defined by the β_1 to β_{1+m} parameters, where $\beta_i = 0$ if t $< K_{i-1}$. The knots are thus reflecting transitions between phases of different slopes in the development of the traits, and at each knot, the slope is changed according to the β of the next segment. Therefore, the times when the knots occur have to be properly defined in order to provide accurate descriptions of the data under investigation, and also their numbers in order to avoid overadaptation to data[59],[28]. We found use of two knots the most suitable for tracheid dimension traits across the time intervals investigated. Hence, the linear spline model used was defined as:

$$y(t) = \beta_0 + \beta_1 t + \beta_2 (t - K_1)_+ + \beta_3 (t - K_2)_+ + \varepsilon_i(t), \qquad \varepsilon_i(t) \overset{i.i.d.}{\sim} N(0, \sigma^2). \qquad [3]$$

In a first analysis, fixed values of K_1 and K_2 were adapted for each trait. Then, the intercept β_0, and the slope parameters β_1, β_2 and β_3 were estimated for each tree by standard least squares[60]. The four estimates were used as the latent trait in the subsequent QTL analysis conducted in R-studio[61], and then analysed using the LASSO model in order to identify SNPs showing significant associations to the traits.

The LASSO model as described by Li *et al* (2014)[59], was applied to all latent traits for the detection of QTLs.

The LASSO model:

$$\min_{(\alpha_0, \alpha_j)} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \alpha_0 - \sum_{j=1}^{p} x_{ij}\alpha_j)^2 + \lambda \sum_{j=1}^{p} |\alpha_j|, \qquad [4]$$

where y_i is the phenotypic value of an individual i ($i=1,...,n$; n is the total number of individuals) for the latent trait, α_0 is the population mean parameter, x_{ij} is the genotypic value of individual i and marker j coded as 0, 1 and 2 for three marker genotypes AA, AB and BB, respectively, α_j is the effect of marker j ($i=1,...,n$; n is the total number of markers), and λ (>0) is a shrinkage tuning parameter. Stability selection probability (SSP) of each SNP was applied as a way to control the false discovery rate and determine significant SNPs[62,63,59]. For a marker to be declared significant, a SSP inclusion ratio (Frequency) was used with an inclusion frequency of at least 0.52 for all traits. This frequency inferred that the expected number of falsely selected markers was less than one (1), according to the formula of Buhlmann *et al,* (2014)[64]. Population structure was accounted for in all analyses by including the first five principal components based on the genotype data as covariates into the model. The LASSO regression has a limitation in that it might over-shrink the effect size of SNPs due to the use of a single tuning parameter for all the regression parameters[65]. The consequence is that the LASSO might significantly under-estimate the proportion of phenotypic variation (PVE) explained by a SNP[66]. To improve this, an adaptive LASSO approach[65] was used alternatively to evaluate the PVE of a QTL (Methods S4):

In brief, estimated breeding values (EBV) were computed for each annual ring by cambial age to reduce site and block effects (see Chen et al 2014). In a second step, linear splines were applied to reconstruct time trajectories based on annual ring EBV. Fix age values for two knots were determined, as the intercept and slope parameters, the latent traits, were fitted to the EBV describing the shape of the time trajectories of each individual tree.

Candidate gene mining

To assess putative functionality of SNPs with significant associations, gene ontology and network analysis of putative genes and their associated orthologs was performed against the NorWood v1.0 database (http://norwood.congenie.org[67]) hosted by ConGenIE (http://congenie.org/). After the identification of the QTL, the Norway spruce contigs linked to the significant SNPs were extracted from the web based database congenie (congenie.org/blast). The complete Norway spruce contigs that harboured the QTLs that were not annotated in the ConGenIE were used to perform a nucleotide BLAST (Blastn) search, using the option for only highly similar sequences (megablast) in the National Center for Biotechnology Information (NCBI) nucleotide collection database (https://blast.ncbi.nlm.nih.gov/Blast.cgi?).

Results and Discussion

Trait trajectories

For traits with complex time/age trajectories, the application of functional mapping enables an aggregated analysis of temporal trends[27]. The ring MFA initially decreased from an average across the trees of about 30° at the pith and stabilized after reaching a cambial age of about 10 years at an average of 10-12°[68]. The adapted central peak curves combined with the threshold at 20° resulted in an average of five years for MFA_{TA}, defining the inner core of lower quality timber with AM performed for the latent traits of MFA_{CORE} and MFA_{OUTER}.

For all the other tracheid phenotypes: wall thickness, radial tracheid width, tangential tracheid width and coarseness, family means of β_0 (intercept) and β_1 to β_3 (effects of knot 1 to 3, see Baison et al., 2019)[69] were implemented in the association mapping. Candidate gene loci were identified for MFA_{CORE}, MFA_{OUTER} and MFA_{TA}, and for the intercepts $\beta_{0,}$ of the tracheid dimensions and coarseness of rings, EW, TW and LW.

Genetic associations detected and modes of gene action

A total of 30 significant associations were detected across the 18 traits with fraction of phenotypic variances being explained (PVE) ranging between 0.01 to 3.79% (Table 2), using an Stability selection probability (SSP) minimum inclusion frequency of 0.52. Seven of the 30 marker trait associations for

which dominance and additive effects could be calculated were consistent with partially to fully

dominant effects ($0.50 < |d/a| < 1.25$). The remaining 23 markers were all determined to have an

additive ($|d/a| < 0.50$) mode of inheritance (Table 2). The relationships between the genotypic classes

of markers associated to a phenotype were consistent with these patterns (Fig. 1). Three SNPs

MA_10436040g0010_171180, MA_105586g0010_65505 and MA_10426383g0010_135796, were

significantly associated *across* and *within* several traits, with all the modes of gene action being

additive for the marker-trait interactions for the three SNPs (Table 2; Fig. 1).
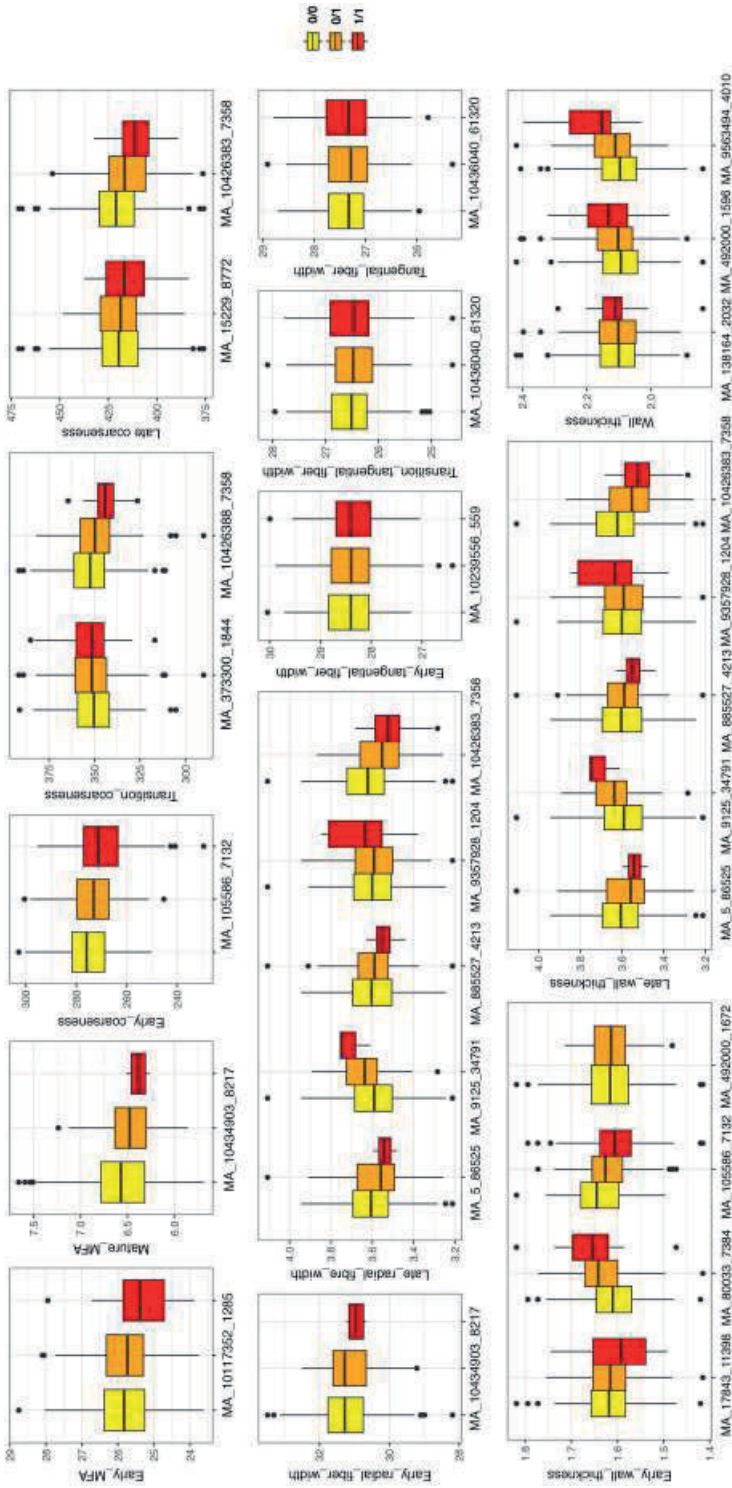
Fig. 1. Box plot of the estimated genotypic effects for all significant associations identified in the study. The middle line represents the median value of the phenotype with that of the genotype. Upper and lower bounds of the box are the 25% (Q1) and 75% (Q3) quantile. Whiskers are Q1-1.5*Interquartile range (IQR) and Q3+1.5*IQR, therefore the outliers are values outside the range (Q1-1.5*IQR to Q3+1.5*IQ. Yellow, orange and red colored boxplots indicate the genotypic classes per SNP. Alleles (0/1) for all the SNPs have been defined in Table 2.

Table 2. Phenotype, QTL position, allele frequencies and modes of allele inheritance.

Phenotype	QTL[1]	SNP[2]	Region, Feature	Alleles (0/1)	Inclusion Frequency	PVE (H^2_{QTL})	$2a^3$	d^4	d/a
MFA$_{CORE}$	129716	MA_10117352g0010_1285	Synonymous	T/C	0.548	1.47%	0.403	0.168	0.836
MFA$_{OUTER}$	165836	MA_10434903g0010_8217	Intron variant	T/A	0.600	1.03%	0.190	-0.002	-0.025
TW$_{EW}$	166535	MA_10435070g0010_17636	Missense	T/C	0.627	3.16%	0.349	-0.053	-0.304
TW$_{LW}$	95509	MA_336364g0010_6123	Missense	C/T	0.665	0.05%	0.007	-0.039	-11.489
	12016	MA_11172g0010_18275	Splice variant	A/G	0.579	0.01%	0.051	-0.078	-3.058
	160388	MA_10433459g0010_6695	Upstream	G/T	0.604	1.72%	0.118	0.042	0.701
	44384	MA_64438g0010_10851	Missense	A/C	0.546	0.03%	0.395	0.024	0.123
	59913	MA_96801g0010_10438	Synonymous	C/T	0.547	0.10%	0.212	0.022	0.211
	116013	MA_950574g0010_7132	Upstream	T/A	0.558	0.05%	0.451	0.004	0.017
	120168	MA_7460525g0010_2105	Upstream	G/A	0.561	2.27%	0.140	0.045	0.642
TW$_{EW}$	131776	MA_10239556g0010_	Upstream	T/G	0.633	1.80%	0.072	0.002	0.069
TW$_{TW}$	**171180**	MA_10436040g0010_61320	Upstream	G/A	0.661	2.13%	0.003	-0.039	-24.638
TW$_{Rring}$	**171180**	MA_10436040g0010_61320	Upstream	G/A	0.566	3.79%	0.019	-0.046	-4.646

15

Group	Position	Gene ID	Effect	Change					
WT$_{EW}$	51296	MA_80033g0010_7384	Missense	T/C	0.541	0.01%	0.047	0.0007	0.033
	65505	MA_105586g0010_7132	Missense	T/C	0.536	0.10%	0.032	-0.002	-0.114
	19482	MA_17843g0010_11398	Missense	A/G	0.546	**0.01%**	0.021	0.010	0.976
	103329	MA_492000g0010_1672	Upstream	A/T	0.582	**0.10%**	0.032	-0.021	-1.277
WT$_{LW}$	1	MA_5g0010_86525	Upstream	G/A	0.876	0.01%	0.046	0.008	0.356
	9848	MA_9125g0010_34791	Upstream	T/C	0.64	0.10%	0.053	0.004	0.131
	112677	MA_885527g0010_4213	Upstream	C/T	0.795	0.02%	0.059	0.018	0.591
	126271	MA_9357928g0010_1204	Upstream	G/A	0.567	0.02%	0.060	-0.024	-0.809
	135796	MA_10426383g0010_7358	Synonymous	A/G	0.61	1.57%	0.097	-0.0007	-0.015
WT$_{Ring}$	103326	MA_492000g0010_1596	Synonymous	T/C	0.726	1.78%	0.038	-0.003	-0.163
	127327	MA_9563494g0010_4010	Missense	G/T	0.509	0.01%	0.084	-0.019	-0.464
	78937	MA_138164g0010_2032	Downstream	C/T	0.529	1.25%	0.007	0.001	0.442
C$_{EW}$	**65505**	MA_105586g0010_7132	Missense	T/C	0.633	2.08%	5.481	-0.238	-0.087
C$_{TW}$	96993	MA_373300g0010_1844	Upstream	C/T	0.559	3.62%	1.905	0.791	0.831
C$_{LW}$	**135796**	MA_10426383g0010_7358	Synonymous	A/G	0.53	3.25%	9.214	0.533	0.116
	16320	MA_15229g0010_	Upstream	A/T	0.512	0.78%	4.120	1.448	0.703
	135796	MA_10426383g0010_7358	Synonymous	A/G	0.635	1.40%	9.434	0.146	0.031

[1]Bold QTLs indicate associations that have been detected *across* and *within* traits.

[2]SNP: The SNP name is composed of contig (MA_number) and SNP position on contig. As an example, the SNP MA_10117352g0010_1285 is located on contig MA_10117352 at position 1285 bp. QTL is the value extracted after the Stability selection probability (SSP) has been performed to indicate the significant associations. This value then points to the significant SNP.

[3] Calculated as the difference between the phenotype means observed within each homozygous class ($2a = |G_{BB}-G_{bb}|$, where G_{ij} is the trait mean in the *ij*th genotype class).

[4] Calculated as the difference between the phenotypic mean observed within the heterozygous class and the average phenotypic mean across both homozygous classes [$d = G_{Bb}-0.5(G_{BB}+G_{bb})$, where G_{ij} is the trait mean in the *ij*th genotypic class]

Genetic associations and genes of interest

Two of the associations detected for MFA were intron variants MA_10434903g0010_8217 and

MA_10117352g0010_1285. MA_10434903g0010_8217 was associated with MFA_{OUTER} explaining

1.03% of the PVE. MA_10117352g0010_1285, a synonymous variant explaining 1.47% PVE

associated with $MFA_{OUTER,}$ occurred within gene MA_10117352g0010 homologous to *Arabidopsis*

ONE HELIX PROTEIN (OHP). The gene is highly expressed especially in needles and shoots in

spruce (Fig. 2). OHPs have been reported to be constitutively expressed and essential for

photosynthesis in Arabidopsis, with mutants exhibiting severe growth defects[70].

Associations for radial tracheid widths were detected in earlywood and latewood. TWr_{EW} was

associated with a single missense SNP (MA_10435070g0010_17636) explaining 3.16% of the PVE

and occurred within a gene encoding nuclear transcription factor Y subunit A-7 (NF-YA7) (Table S1).

NF-Y is a multimer complex binding CCAAT box in the promoter regions of many genes, and has

multiple biological functions including growth regulation, cell size regulation, and responses to abiotic

stresses[71,72], including nitrogen deficiency in *Arabidopsis*[73]. The overexpression of the NF-YAs has

been shown to stimulate growth during low nitrogen and phosphorous availability[74]. This gene is

ubiquitously, highly expressed in shoots and buds of spruce, indicating its important function in this

species (Fig. 2

Fig 2. The heatmap showing levels of the variance stabilized transformed expression values (VST values) of spruce candidate genes in different organs and tissues based on ExAtlas data[75] available at http://congenie.org.

TWr$_{LW}$ with seven significant associations, had the highest number of detected associations per trait. Two missense SNPs, MA_336364g0010_6123 and MA_64438g0010_10851 associated with TWr$_{LW}$, explained a small proportion of the PVE observed 0.01% and 0.03%, respectively. MA_336364g0010 is homologous to the *Arabidopsis INDUCER OF CBF EXPRESSION 2 (ICE2)* regulating deep-freezing tolerance by inducing *CBF1, CBF2* and *CBF3* genes (Table S1)[76]. *CBF* genes have been identified to constitute a central node of hormone cross-talk during cold stress response and their expression is modulated by abscisic acid, gibberelins, jasmonate, ethylene and brassinosteroids[77]. It has emerged that different hormone signaling pathways converge at the *CBF* promoter level, with the result of this hormone cross-talk being the fine-tuned transcript levels impacting on plant development and growth[78]. In spruce, the homolog of *ICE2* gene is highly expressed in developing stems (Fig. 2) and strongly upregulated in the cambium and radial expansion zone (Fig. 3) supporting its role *in situ* in promoting the tracheid expansion. Since *CBF*s have already been identified as convergence points for hormones required for the regulation of plant growth under cold stress, these factors would warrant a detailed look in relation to their influence on wood tracheid development, especially during the time when the water stress and cold stress can be common. The gene MA_64438g0010 is homologous to an *Arabidopsis PHOSPHATIDYLINOSITOL BINDING CLATHRIN ASSEMBLY PROTEIN 5B (PICALM5B),* a part of the ENTH/ANTH/VHS superfamily (Table S1). The ENTH/ANTH/VHS superfamily is involved in clatrin assembly at secretory vescicles and is essential for vesicle intracellular trafficking and thus, cell growth and development[79]. The gene was observed expressed in developing wood (Fig. 3), indicating its importance for tracheid development in spruce. Indeed, the genes of ENTH/ANTH/VHS family have been previously associated with secondary cell wall formation in *Populus*[80], and vescicle trafficking-related genes were seen upregulated coinciding with radial expansion of developing wood cells in aspen[81]. Such genes are therefore expected to be associated with tracheid radial expansion in spruce. Another gene associated with TWr$_{LW}$ was MA_950574g0010_7132, explaining a comparatively high PVE of 2.27%. It is remotely similar to *Arabidopsis CALCINEURIN-B-LIKE-INTERACTING SERINE/THREONINE-PROTEIN KINASE 23 (CBLPK23)* involved in the regulation of HAK5-mediated high-affinity K$^+$

uptake in calcium-dependent manner in *Arabidopsis* roots[82]. The confidence of the spruce model was low, but the gene was found highly expressed in developing shoots, buds and cones (Fig. 2), and during primary and secondary wall formation in developing spruce tracheids (Fig. 3) confirming that it was not a pseudogene. A *CALCINEURIN-B-LIKE* gene was found to explain the largest phenotypic variance in cell wall mannose content in white spruce[23]. These observations make the identified spruce *CBLPK23* gene an interesting candidate for calcium-dependent regulation of K^+ uptake in developing tracheids, thus likely regulating tracheid expansion, similar to vessel element expansion, known to be dependent on K^+ transport[83]. Interestingly, there was another candidate gene related to K^+ transport associated with tracheid radial width: the splice variant MA_11172g0010_18275 explaining 0.01% PVE (Table 2). This gene is homologous to *Arabidopsis CYCLIC NUCLEOTIDE-GATED CHANNEL 17* (*CNGC17*) (Table S1). CNGCs are potassium channels involved in several plant physiological processes including root development, pollen tube growth and plant disease resistance[84]. They regulate ion homeostasis within plants through the uptake of cations, which is essential for plant growth and development[85]. *Arabidopsis* CNGC17 is localized in the plasmamembrane and promotes protoplast expansion by regulating cation uptake[86]. Its spruce homolog exhibited specific expression during latewood formation in August (Fig. 2), supporting its role in latewood tracheid development.

Three significant associations were identified for tangential tracheid width components with an upstream variant MA_10436040g0010_61320 being detected *across* traits TWt_{TW} and TWt_{Ring} (Table 2). This variant was detected on contig MA_10436040 with high inclusion frequencies explaining relatively high percentages of the variance observed, 2.13% for TWt_{TW} and 3.79% for TWt_{Rin} (Table 2). The associated gene - MA_10436040g0010 - is homologous to the stress-related eukaryotic initiation factor 4A-III (eIF4A-III) which also has a DEAD-box ATP-dependent RNA helicase 2, and is involved in RNA processing and nonsense-mediated mRNA decay in *Arabidopsis*, especially under hypoxia and heat stress[87] (Table S1). The spruce gene was not found expressed in available datasets (Fig 2). SNP MA_10239556g0010_131776 was associated with TWt_{EW} and explained a moderate amount of the PVE 1.80% (Table 2). The *Arabidopsis* homolog encodes a subunit C of the vacuolar ATP synthase, which is a membrane-bound enzyme complex/ion transporter that combines ATP synthesis and/or hydrolysis with the transport of protons across the

tonoplast membrane. This gene was highly and ubiquitously expressed (Fig. 2). All three SNPs were consistent with an additive mode of gene action (Table 2).

Twelve associations were detected for wall thickness components, with low to moderate PVE ranging from 0.01 to 1.78% (Table 2). Two of these associations (SNP MA_105586g0010_7132 and MA_10426383g0010_7358) were shared *across* cell wall thickness and coarseness traits. Ring average for cell Wall Thickness (WT_{Ring}) had three significant associations. The synonymous SNP MA_492000g0010_1672 had a high inclusion frequency (0.726) and explained the highest percentage of PVE (1.78%). The same SNP was associated with WT_{EW}. The gene *MA_492000g0010* is homologous to a tRNA synthetase beta subunit family protein, phenylalanyl-tRNA synthetase beta chain (Table S1). Consistent with its predicted general metabolic function in protein biosynthesis, it is ubiquitous and highly expressed in spruce tissues (Fig. 2). Missense SNP MA_9563494g0010_4010 and downstream variant MA_138164g0010_2032 explained 0.01% and 1.25% PVE, respectively. MA_9563494g0010_4010 is located in a gene *MA_9563494g0010* named as *Picea abies BIG GRAIN 2* (*PabBG2*)[87] homologous to the *BIG GRAIN 1* gene (*OsaBG1*) in rice[89]. *OsaBG1* encodes a membrane protein regulating auxin transport and sensitivity, and positively affecting plant biomass and seed size. The gene belongs to a small family containing nine members in spruce[88]. Auxin has long been known to act as a key hormone essential for the induction of vascular strands, cambial growth and secondary wall deposition[90,91,92,93]. *PabBG2* is highly expressed and specifically upregulated in the developing xylem (Fig. 2) with a peak of expression in the cambial zone (Fig. 3), coinciding with a peak of IAA distribution in wood forming tissues[91,94]. It is therefore likely that the *PabBG2* gene pays a major role in xylogenesis, as suggested by its association with tracheid cell wall thickness, and that it should be considered as main target for woody biomass increase. Moreover, the SNP MA_138164g0010_78937 explaining PVE 1.25% associated with WT_{Ring} was located in a gene homologous to the subunit of E3 ubiquitin complex encoded by *AtAPC1* and involved in cell cycle regulation by degradation of cyclin B1[95]. The E3 ubiquitin complex is also known in *Arabidopsis* to regulate auxin homeostasis[96,97,97]. Hence, the detection of two significant associations for WT_{Ring} that are potentially related to auxin regulation implies a close relation between auxin and cell wall thickness in spruce. A QTL in rice grain for width and weight, which is related to plant biomass, has

been associated with a RING-type E3 ubiquitin ligase[99]. Several auxin responsive genes were also associated with tracheid width and MFA, which both are linked to cell wall thickness, in white spruce[23].

WT$_{EW}$ has three significant associations beside MA_492000g0010_103329 discussed above (Table 2). The missense variant MA_80033g0010_51296 was within a gene encoding a MYB transcription factor similar to *Arabidopsis* MYB68 (Table S1). This gene exhibited very low expression levels in the developing xylem but rather was expressed in young shoots and needles (Fig. 3). Different MYB transcription factors regulate plant developmental processes, and several have been identified as crucial factors for secondary wall deposition and lignification. Loblolly pine (*Pinus teada* L.) *PtMYB8* expressed in spruce induced secondary cell wall thickening[100]. White spruce (*P. glauca* L.) *PgMYB4* was associated with cell wall thickness and tracheid coarseness[23], and has been shown to be highly expressed during secondary cell wall formation and lignification in both white spruce and loblolly pine[101]. MYB encoded by *MA_80033g0010* could play a more indirect role in secondary wall regulation in spruce considering its expression (Fig. 2). Two remaining SNPs MA_17843g0010_19482 and MA_105586g0010_7132, had PVEs of 0.01% and 0.10%, respectively (Table 2). The former was a missense variant within a gene homologous to *Arabidopsis TOC64-V*. The latter was not matching any known gene and was also associated with C$_{EW}$ and explaining a moderate percentage of PVE 2.08%. However, the two models were not expressed in any of the previously reported spruce expression studies (Fig. 2).

WT$_{LW}$ was associated with four upstream variants and a single synonymous SNP MA_10426383g0010_7358. The four upstream variants explained PVE ranging from 0.01 to 0.10% whereas the synonymous SNP MA_10426383g0010_7358 had a high inclusion frequency and explained a moderate amount of the PVE 1.57% (Table 2). *MA_10426383g0010* is homologous to *VIT_16s0098g01810* from *Vitis vinifera* (Table S1) annotated as encoding ATP binding protein that may be involved in chromosome organization and biogenesis[101]. The *Arabidopsis* homolog - *GAMMA-IRRADIATION AND MITOMYCIN C INDUCED 1* (*GMI1*) is responsible for double strand repair via somatic homologous recombination[103]. The spruce gene shows increased expression in organs with active meristems (Fig. 2), which is expected for the function in DNA repair. The same SNP

MA_10426383g0010_7358 was also associated with traits related to coarseness (C_{TW} and C_{LW}) and explained a relatively high PVE of 3.25% and 1.40%, respectively. It also had high inclusion frequencies for all three traits (WT_{LW}, C_{TW} and C_{LW}) (Table 2). The associated gene might therefore be a good candidate to explore for effects on tracheid development, especially since it is highly expressed in the developing wood[75] (Fig. 2). SNP MA_5g0010_1 associated with WT_{LW} was detected upstream of gene *MA_5g0010* belonging to the 4-coumarate-CoA ligase (4CL) family, which includes key enzymes in the monolignol biosynthetic pathway. However, the *Arabidopsis* homolog of *MA_5g0010, At4g05160* does not encode an enzyme active on phenyl propanoid substrates but a fatty acyl CoA synthase involved in lipid and jasmonic acid biosynthesis[104]. *MA_5g0010* is not expressed in the developing wood but it is highly expressed in young vegetative shoots and needles, including the infected needles (Fig. 2), making it an unlikely candidate for lignin biosynthesis in developing wood but suggesting a rather indirect function in the regulation of tracheid cell wall thickness. The SNP MA_9125g0010_34791 associated with WT_{LW} was located upstream of a gene homologous to *Arabidopsis OBERON2 (OBE2)* encoding a plant homeodomain (PHD) finger protein (Table S1) (Lee *et al.*, 2009). Homeodomain genes encode transcription factors central in the regulation of plant developmental processes[105]. *OBE1* and *OBE2* redundanlty regulate meristem establishment and maintenance in *Arabidopsis* (Saiga *et al.*, 2008). The spruce *OBE2* gene is ubiquitous and highly expressed in vegetative and reproductive organs (Fig. 2) including developing wood where it shows high expression during secondary wall deposition (Fig. 3) and therefore it could have a direct role in cell wall thickening in tracheids. SNP MA_885527g0010_112677 associated with WT_{LW} was found upstream of a gene containing a SET domain. SET domain proteins have been identified in *Arabidopsis* to play aide in the epigenetic control of genes involved in a wide range of activities including plant growth[100]. A link has also been established between PHD finger proteins and SET domain proteins in the regulation of developmental transitions in *Arabidopsis* where PHD finger proteins VEL1, VRN5 and VIN3 interacting with H3K27me3 repress FLC transcription allowing for the transition from vegetative to reproductive development[106]. *MA_885527g0010* is highly upregulated in developing wood from August that is involved in latewood biosynthesis (Fig. 2) suggesting its direct role in latewood tracheid development.
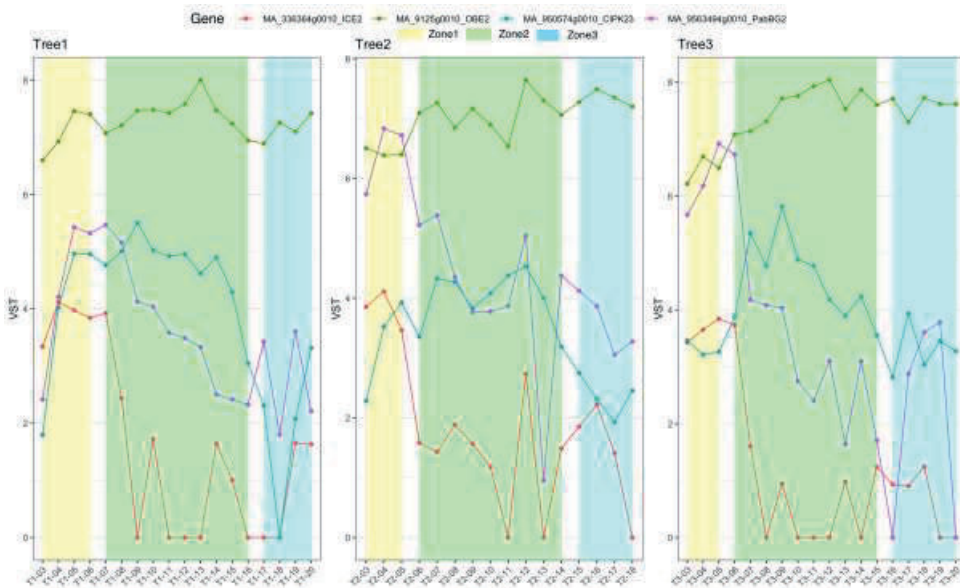
Fig. 3. Expression (Variance stabilized transformed expression values) profiles of selected candidate genes in wood developing tissues of sections through developing wood zones, phloem to mature xylem of spruce based on NorWood dataset (http://norwood.congenie.org/norwood-v1.0/[67]). Expression profiles of three trees sampled during the peak of wood formation in the summer are shown. The X-axis shows numbers of consecutive tangential sections through the developing wood zones. The zone numbers corresponding to: i) cambium-radial expansion zone,ii) secondary wall formation zone, and iii) mature zone are shown above the graphs for each tree.

A total of five significant associations were identified for coarseness traits explaining moderate to high PVE ranging from 0.78 to 3.62% (Table 2). Two of them, SNPs MA_105586g0010_7132 and MA_10426383g0010_7358 were also associated with WT_{EW} and WT_{LW}, and discussed above. An Upstream variant MA_373300g0010_1844 associated with C_{TW} explained a relatively high percentage of PVE 3.62% and was consistent with a partial to fully dominant mode of gene action (Table 2) as shown by the genotypic effects (Fig 1). The gene is similar to *Potri.T064000* from *Populus trichocarpa* annotated as encoding a protein kinase similar to wall-associated receptor

kinase-like (WAK-like) proteins. WAKs have been previously reported to be associated with average ring width and the proportion of earlywood in white spruce[23] and with MFA in *Populus*[80]. The gene is expressed primarily in early season needles and late season stem from vegetative shoots but there is no detectable expression in developing xylem (Fig. 2), suggesting its indirect involvement in the regulation of tracheid coarseness.

Conclusion

This work presents the first genome wide dissection of wood tracheid traits in Norway spruce. A total of 30 significant associations were detected for all investigated traits. These associations have identified a set of genes that could be exploited to alter wood tracheid traits for improving solid wood properties for its use in industrial processes. Previous studies utilizing a LASSO penalized analysis approach were limited in the nature and number of molecular markers available[107,27], with our study representing a major advance by using 178101 SNPs with a functional mapping approach. The relatively small number of associations is comparable to other association studies of complex growth traits in forest trees, were a few associations are detected with a relatively small proportion of the genetic variation being explained[108,80,24,109,110]. It can be argued that many of the alleles causing variation for polygenic traits may be either rare or have small effects and current GWAS methods lack the power to detect them, thus the small number of significant associations[111,112]. The small number of associations being reported could also be largely due to the small sample sizes in these studies for such complex traits. Theoretical work has also shown that alleles of large effect are unusual, with allele effect having been suggested to follow a negative exponential distribution pattern[113]. Thus the magnitude of the detected allele effects follows a truncated exponential distribution[114]. Therefore, the detection of alleles with small effects is difficult when compounded with the small population size. The small number of significant associations can also be attributed to the genotyping method, which is a complexitiy reduction genotyping method. The limitation of the genotyping employed in our study has also been noted in other studies[115], in that some of the alleles impacting a trait might not be within the captured regions that we targeted. If the sampled markers do not include the casual allele or if the

LD between the marker and the casual allele is incomplete the power of detection is drastically reduced[80]. The statistical power required to detect associations between molecular markers and a trait is heavily dependent upon the sample size[116,117]. Due to the challenges of developing large populations for GWAS in conifers, the majority of the studies utilize a few hundred individuals from natural populations, which limits the statistical power of GWAS. It was reported that in order to capture 50% of genetic variaon for growth traits in an association mapping study, it would require roughly 25 000 individuals to be analysied[118]. Therefore, the relatively small association population size results in low statistical power, thus rendering small to medium effect QTLs statistically non-significant and very difficult to detect. Our study had 517 martenal trees to perform GWAS upon, thus rendering a small number of significant associations. Missing heritability will remain an issue in association studies as long as population sizes are kept in the range of hundreds[118]. However, improvements made to statitstical methods are now potential viable options, which are being developed and utilize a combination of information from multiple populations using Meta-GWAS and Joint-GWAS[119,120]. These approaches are now being applied in some recent forest tree studies[117] and could be the next level of analysis using our application of latent traits on these complex traits.

References

1. Mäkinen, H., Saranpää, P. & Linder, S. Effect of Growth Rate on Fibre Characteristics in Norway Spruce (Picea abies (L.) Karst.). (2002).

2. Lundqvist, S.-O. & Gardiner, B. Key products of the forest-based industries and their demands on wood raw material properties. *Joensuu, Finland, http://www. efi. int/files/attachments/publications/eforwood/efi_tr_71. pdf,[March 23, 2013]* (2011).

3. Brändström, J. *Morphology of Norway spruce tracheids with emphasis on cell wall organisation.* vol. 237 (2002).

4. Dutilleul, P., Herman, M. & Avella-Shaw, T. Growth rate effects on correlations among ring width, wood density, and mean tracheid length in Norway spruce (Picea abies). *Canadian Journal of Forest Research* **28**, 56–68 (1998).

5. Hannrup, B. & Ekberg, I. Age-age correlations for tracheid length and wood density in Pinus sylvestris. *Can. J. For. Res.* **28**, 1373–1379 (1998).

6. Lindström, H. Fiber length, tracheid diameter, and latewood percentage in Norway spruce: development from pith outward. *Wood and Fiber Science* **29**, 21–34 (2007).

7. Lundqvist, S.-O., Grahn, T. & Hedenberg, Ö. Models for fibre dimensions in different softwood species. Simulation and comparison of within and between tree variations for Norway and Sitka spruce, Scots and Loblolly pine. in vol. 5 22–27 (2005).

8. Hannrup, B. *et al.* Genetic parameters of growth and wood quality traits in Picea abies. *Scandinavian Journal of Forest Research* **19**, 14–29 (2004).

9. Yang, J. L. & Evans, R. Prediction of MOE of eucalypt wood from microfibril angle and density. *Holz als Roh- und Werkstoff* **61**, 449–452 (2003).

10. Chen, Z.-Q. *et al.* Inheritance of growth and solid wood quality traits in a large Norway spruce population tested at two locations in southern Sweden. *Tree Genetics & Genomes* **10**, 1291–1303 (2014).

11. Eckert, A. J. *et al.* High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (Pinus taeda L.). *Tree Genetics & Genomes* **5**, 225–234 (2009).

12. Neale, D. B. & Ingvarsson, P. K. Population, quantitative and comparative genomics of adaptation in forest trees. *Current Opinion in Plant Biology* **11**, 149–155 (2008).

13. Parchman, T. L. *et al.* Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* **21**, 2991–3005 (2012).

14. Hall, D., Hallingbäck, H. R. & Wu, H. X. Estimation of number and size of QTL effects in forest tree traits. *Tree Genetics & Genomes* **12**, 110 (2016).

15. Hall, D., Tegström, C. & Ingvarsson, P. K. Using association mapping to dissect the genetic basis of complex traits in plants. *Briefings in Functional Genomics* elp048 (2010).

16.   Neale, D. B. & Savolainen, O. Association genetics of complex traits in conifers. *Trends in Plant Science* **9**, 325–330 (2004).

17.   Cardon, L. R. & Bell, J. I. Association study designs for complex diseases. *Nature Reviews Genetics* **2**, 91 (2001).

18.   Pavy, N., Namroud, M. C., Gagnon, F., Isabel, N. & Bousquet, J. The heterogeneous levels of linkage disequilibrium in white spruce genes and comparative analysis with other conifers. *Heredity* **108**, 273–284 (2012).

19.   Thavamanikumar, S. *et al.* Association mapping for wood quality and growth traits in Eucalyptus globulus ssp. globulus Labill identifies nine stable marker-trait associations for seven traits. *Tree Genetics & Genomes* **10**, 1661–1678 (2014).

20.   Larsson, H., Källman, T., Gyllenstrand, N. & Lascoux, M. Distribution of Long-Range Linkage Disequilibrium and Tajima's D Values in Scandinavian Populations of Norway Spruce (Picea abies). *G3: Genes|Genomes|Genetics* **3**, 795–806 (2013).

21.   Gupta, P. K., Rustgi, S. & Kulwal, P. L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Molecular Biology* **57**, 461–485 (2005).

22.   Baison, J. *et al.* Genome-wide association study identified novel candidate loci affecting wood formation in Norway spruce. *The Plant Journal* **100**, 83–100 (2019).

23.   Beaulieu, J. *et al.* Association Genetics of Wood Physical Traits in the Conifer White Spruce and Relationships With Gene Expression. *Genetics* **188**, 197–214 (2011).

24.   McKown, A. D. *et al.* Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of Populus trichocarpa. *New Phytologist* **203**, 535–553 (2014).

25.   Ma, C.-X., Casella, G. & Wu, R. Functional Mapping of Quantitative Trait Loci Underlying the Character Process: A Theoretical Framework. *Genetics* **161**, 1751–1762 (2002).

26.   Xing, J. U. N., Li, J., Yang, R., Zhou, X. & Xu, S. Bayesian B-spline mapping for dynamic quantitative traits. *Genetics Research* **94**, 85–95 (2012).

27.   Li, Z. *et al.* Functional multi-locus QTL mapping of temporal trends in Scots pine wood traits. *G3: Genes| Genomes| Genetics* **4**, 2365–2379 (2014).

28.   Camargo, A. V. *et al.* Functional Mapping of Quantitative Trait Loci (QTLs) Associated With Plant Performance in a Wheat MAGIC Mapping Population. *Frontiers in Plant Science* **9**, 887–887 (2018).

29.   Via, B. K., Stine, M., Shupe, T. F., So, C.-L. & Groom, L. Genetic Improvement of Fiber Length and Coarseness Based on Paper Product Performance and Material Variability – a Review. *IAWA Journal* **25**, (2004).

30.   Capron, A. *et al.* Identification of quantitative trait loci controlling fibre length and lignin content in Arabidopsis thaliana stems. *J Exp Bot* **64**, 185–197 (2013).

31.   Thamarus, K. *et al.* Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of Eucalyptus globulus. *Theoretical and Applied Genetics* **109**, 856–864 (2004).

32.   Sewell, M. *et al.* Identification of QTLs influencing wood property traits in loblolly pine (Pinus taeda L.). II. Chemical wood properties. *Theoretical and Applied Genetics* **104**, 214–222 (2002).

33.   Thumma, B. R. *et al.* Quantitative trait locus (QTL) analysis of wood quality traits in Eucalyptus nitens. *Tree Genetics & Genomes* **6**, 305–317 (2010).

34.   Li, Z. *et al.* Functional multi-locus QTL mapping of temporal trends in scots pine wood traits. *G3: Genes, Genomes, Genetics* **4**, 2365–2379 (2014).

35.   Mellerowicz, E. J., Baucher, M., Sundberg, B. & Boerjan, W. Unravelling cell wall formation in the woody dicot stem. in *Plant Cell Walls* 239–274 (Springer, 2001).

36.   Plomion, C., Leprovost, G. & Stokes, A. Wood formation in trees. *Plant Physiology* **127**, 1513–1523 (2001).

37.   Masuda, Y. Auxin-induced cell elongation and cell wall changes. *The botanical magazine = Shokubutsu-gaku-zasshi* **103**, 345 (1990).

38.   Farquharson, K. L. Probing the Role of Auxin in Wood Formation. *The Plant Cell* **20**, 822–822 (2008).

39.   Wegrzyn, J. L. *et al.* Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (Populus trichocarpa, Salicaceae) secondary xylem. *New Phytologist* **188**, 515–532 (2010).

40.   Buell, C. R., Hirsch, C. N., Hirsch, C. D. & Evans, J. Reduced representation approaches to interrogate genome diversity in large repetitive plant genomes. *Briefings in Functional Genomics* **13**, 257–267 (2014).

41.   Chen, Z.-Q. *et al.* Inheritance of growth and solid wood quality traits in a large Norway spruce population tested at two locations in southern Sweden. *Tree Genetics & Genomes* **10**, 1291–1303 (2014).

42.   Hayatgheibi, H. *et al.* Genetic control of transition from juvenile to mature wood with respect to microfibril angle in Norway spruce (Picea abies) and lodgepole pine (Pinus contorta). *Canadian Journal of Forest Research* **48**, 1358–1365 (2018).

43.   Evans, R. & Downes, G. M. *Recent developments in automated wood quality assessment*. (CRC Publications Committee, 1994).

44.   Lundqvist, S.-O. *et al.* Age and weather effects on between and within ring variations of number, width and coarseness of tracheids and radial growth of young Norway spruce. *European Journal of Forest Research* **137**, 719–743 (2018).

45.   Kostiainen, K. *et al.* Stem wood properties of mature Norway spruce after 3 years of continuous exposure to elevated [CO2] and temperature. *Global Change Biology* **15**, 368–379 (2009).

46.   Franceschini, T. *et al.* Empirical models for radial and tangential fibre width in tree rings of Norway spruce in north-western Europe. *Holzforschung* **66**, 219–230 (2012).

47.   Fries, A., Ulvcrona, T., Wu, H. X. & Kroon, J. Stem damage of lodgepole pine clonal cuttings in relation to wood and fiber traits, acoustic velocity, and spiral grain. *Scandinavian Journal of Forest Research* **29**, 764–776 (2014).

48.   Chen, Z.-Q. *et al.* Inheritance of growth and solid wood quality traits in a large Norway spruce population tested at two locations in southern Sweden. *Tree Genetics & Genomes* **10**, 1291–1303 (2014).

49.   Chen, Z.-Q. *et al.* Genetic analysis of fiber dimensions and their correlation with stem diameter and solid-wood properties in Norway spruce. *Tree Genetics & Genomes* **12**, 123 (2016).

50.   Zhou, L. *et al.* Effect of number of annual rings and tree ages on genomic predictive ability for solid wood properties of Norway spruce. *BMC genomics* **21**, 1–12 (2020).

51.   Lundqvist, S.-O. *et al.* Age and weather effects on between and within ring variations of number, width and coarseness of tracheids and radial growth of young Norway spruce. *European Journal of Forest Research* **137**, 719–743 (2018).

52.   Hayatgheibi, H. *et al.* Genetic control of transition from juvenile to mature wood with respect to microfibril angle in Norway spruce (Picea abies) and lodgepole pine (Pinus contorta). *Canadian Journal of Forest Research* **48**, 1358–1365 (2018).

53.   Vidalis, A. *et al.* Design and evaluation of a large sequence-capture probe set and associated SNPs for diploid and haploid samples of Norway spruce (Picea abies). *bioRxiv* (2018) doi:10.1101/291716.

54.   Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

55.   Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

56.   McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).

57.   Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

58.   Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).

59.   Li, Z. & Sillanpää, M. J. A Bayesian Nonparametric Approach for Mapping Dynamic Quantitative Traits. *Genetics* **194**, 997–1016 (2013).

60.   Ruppert, D., Wand, M. P. & Carroll, R. J. *Semiparametric regression*. vol. 12 (Cambridge university press, 2003).

61.   Team, Rs. RStudio: integrated development for R. *RStudio, Inc., Boston, MA URL http://www. rstudio. com* (2015).

62.   Li, H. *et al.* Forward LASSO analysis for high-order interactions in genome-wide association study. *Briefings in Bioinformatics* **15**, 552–561 (2013).

63.   Gao, H. *et al.* Forward LASSO analysis for high-order interactions in genome-wide association study. *Briefings in Bioinformatics* **15**, 552–561 (2014).

64.   Bühlmann, P., Kalisch, M. & Meier, L. High-dimensional statistics with a view toward applications in biology. (2014).

65.   Zou, H. The adaptive lasso and its oracle properties. *Journal of the American statistical association* **101**, 1418–1429 (2006).

66.   Li, Z. & Sillanpää, M. J. Dynamic Quantitative Trait Locus Analysis of Plant Phenomic Data. *Trends in Plant Science* **20**, 822–833 (2015).

67.   Jokipii-Lukkari, S. *et al.* NorWood: a gene expression resource for evo-devo studies of conifer wood development. *New Phytologist* **216**, 482–494 (2017).

68.   Hayatgheibi, H. *et al.* Genetic control of transition from juvenile to mature wood with respect to microfibril angle (MFA) in Norway spruce (Picea abies) and lodgepole pine (Pinus contorta). *bioRxiv* (2018) doi:10.1101/298117.

69.   Baison, J. *et al.* Genome-wide association study identified novel candidate loci affecting wood formation in Norway spruce. *The Plant Journal* **100**, 83–100 (2019).

70.   Beck, J. *et al.* Small One-Helix Proteins Are Essential for Photosynthesis in Arabidopsis. *Frontiers in Plant Science* **8**, (2017).

71.   Zanetti, M. E., Rípodas, C. & Niebel, A. Plant NF-Y transcription factors: Key players in plant-microbe interactions, root development and adaptation to stress. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1860**, 645–654 (2017).

72.   Zhao, H. *et al.* The Arabidopsis thaliana Nuclear Factor Y Transcription Factors. *Frontiers in Plant Science* **7**, (2017).

73.   Sorin, C. *et al.* A miR169 isoform regulates specific NF-YA targets and root architecture in Arabidopsis. *New Phytologist* **202**, 1197–1211 (2014).

74.   Qu, B. *et al.* A Wheat CCAAT Box-Binding Transcription Factor Increases the Grain Yield of Wheat with Less Fertilizer Input. *Plant Physiology* **167**, 411–423 (2015).

75.   Nystedt, B. *et al.* The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579 (2013).

76.   Kim, Y. S., Lee, M., Lee, J.-H., Lee, H.-J. & Park, C.-M. The unified ICE–CBF pathway provides a transcriptional feedback control of freezing tolerance during cold acclimation in Arabidopsis. *Plant Molecular Biology* **89**, 187–201 (2015).

77.   Barrero-Gil, J. & Salinas, J. CBFs at the Crossroads of Plant Hormone Signaling in Cold Stress Response. *Molecular Plant* **10**, 542–544 (2017).

78.   Achard, P. *et al.* The Cold-Inducible CBF1 Factor–Dependent Signaling Pathway Modulates the Accumulation of the Growth-Repressing DELLA Proteins via Its Effect on Gibberellin Metabolism. *The Plant Cell* **20**, 2117–2129 (2008).

79.   De Craene, J.-O. *et al.* Evolutionary analysis of the ENTH/ANTH/VHS protein superfamily reveals a coevolution between membrane trafficking and metabolism. *BMC Genomics* **13**, 297 (2012).

80.   Porth, I. *et al.* Genome-wide association mapping for wood characteristics in Populus identifies an array of candidate single nucleotide polymorphisms. *New Phytologist* **200**, 710–726 (2013).

81.   Sundell, D. *et al.* AspWood: High-spatial-resolution transcriptome profiles reveal uncharacterized modularity of wood formation in Populus tremula. *The Plant Cell* (2017).

82.   Ragel, P. *et al.* CIPK23 regulates HAK5-mediated high-affinity K+ uptake in Arabidopsis roots. *Plant Physiology* 01401.2015 (2015).

83.   Langer, K. *et al.* Poplar potassium transporters capable of controlling K+ homeostasis and K+-dependent xylogenesis. *The Plant Journal* **32**, 997–1009 (2002).

84.   Ma, W., Smigel, A., Verma, R. & Berkowitz, G. A. Cyclic nucleotide gated channels and related signaling components in plant innate immunity. *Plant Signaling & Behavior* **4**, 277–282 (2009).

85.   Kaplan, B., Sherman, T. & Fromm, H. Cyclic nucleotide-gated channels in plants. *FEBS Letters* **581**, 2237–2246 (2007).

86.   Ladwig, F. *et al.* Phytosulfokine regulates growth in Arabidopsis through a response module at the plasma membrane that includes CYCLIC NUCLEOTIDE-GATED CHANNEL17, H+-ATPase, and BAK1. *The Plant Cell* **27**, 1718–1729 (2015).

87.   Pascuan, C., Frare, R., Alleva, K., Ayub, N. D. & Soto, G. mRNA biogenesis-related helicase eIF4AIII from Arabidopsis thaliana is an important factor for abiotic stress adaptation. *Plant cell reports* **35**, 1205–1208 (2016).

88.   Mishra, B. S., Jamsheer, K., Singh, D., Sharma, M. & Laxmi, A. Genome-Wide Identification and Expression, Protein–Protein Interaction and Evolutionary Analysis of the Seed Plant-Specific BIG GRAIN and BIG GRAIN LIKE Gene Family. *Frontiers in Plant Science* **8**, 1812 (2017).

89.   Liu, L. *et al.* Activation of Big Grain1 significantly improves grain size by regulating auxin transport in rice. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 11102–11107 (2015).

90.   Uggla, C., Mellerowicz, E. J. & Sundberg, B. Indole-3-Acetic Acid Controls Cambial Growth in Scots Pine by Positional Signaling. *Plant Physiology* **117**, 113–121 (1998).

91.   Tuominen, H. *et al.* Cambial-Region-Specific Expression of the Agrobacterium iaa Genes in Transgenic Aspen Visualized by a LinkeduidA Reporter Gene. *Plant Physiology* **123**, 531–542 (2000).

92.   Ranocha, P. *et al.* Arabidopsis WAT1 is a vacuolar auxin transport facilitator required for auxin homoeostasis. *Nature communications* **4**, 2625 (2013).

93.   Yang, J. H. & Wang, H. Molecular Mechanisms for Vascular Development and Secondary Cell Wall Formation. *Frontiers in Plant Science* **7**, (2016).

94.   Hellgren, J. M. *Ethylene and auxin in the control of wood formation*. vol. 268 (2003).

95.   Guo, L., Jiang, L., Lu, X.-L. & Liu, C.-M. ANAPHASE PROMOTING COMPLEX/CYCLOSOME-mediated cyclin B1 degradation is critical for cell cycle synchronization in syncytial endosperms. *Journal of integrative plant biology* (2018).

96.   Gray, W. M. *et al.* Identification of an SCF ubiquitin–ligase complex required for auxin response in Arabidopsis thaliana. *Genes & development* **13**, 1678–1691 (1999).

97.   Kepinski, S. & Leyser, O. Auxin-induced SCFTIR1–Aux/IAA interaction involves stable modification of the SCFTIR1 complex. *Proceedings of the National Academy of Sciences* **101**, 12381–12386 (2004).

98.   Azpeitia, E. & Alvarez-Buylla, E. R. A complex systems approach to Arabidopsis root stem-cell niche developmental mechanisms: from molecules, to networks, to morphogenesis. *Plant Molecular Biology* **80**, 351–363 (2012).

99.   Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z. & Lin, H.-X. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nature Genetics* **39**, 623 (2007).

100.   Bomal, C. *et al.* Involvement of Pinus taeda MYB1 and MYB8 in phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative in planta analysis. *Journal of experimental botany* **59**, 3925–3939 (2008).

101.   Bedon, F., Grima-Pettenati, J. & Mackay, J. Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (Picea glauca). *BMC Plant Biology* **7**, 17–17 (2007).

102.   Davies, T. G. E. & Coleman, J. O. D. The Arabidopsis thaliana ATP-binding cassette proteins: an emerging superfamily. *Plant, Cell & Environment* **23**, 431–443 (2000).

103.   Böhmdorfer, G. *et al.* GMI1, a structural-maintenance-of-chromosomes-hinge domain-containing protein, is involved in somatic homologous recombination in Arabidopsis. *The Plant Journal* **67**, 420–433 (2011).

104.   Schneider, K. *et al.* A new type of peroxisomal acyl-coenzyme A synthetase from Arabidopsis thaliana has the catalytic capacity to activate biosynthetic precursors of jasmonic acid. *Journal of Biological Chemistry* **280**, 13962–13972 (2005).

105.   Chew, W., Hrmova, M. & Lopato, S. Role of Homeodomain Leucine Zipper (HD-Zip) IV Transcription Factors in Plant Development and Plant Protection from Deleterious Environmental Factors. *International Journal of Molecular Sciences* **14**, 8122–8147 (2013).

106.   Thorstensen, T., Grini, P. E. & Aalen, R. B. SET domain proteins in plant development. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1809**, 407–420 (2011).

107.   Ma, C.-X., Casella, G. & Wu, R. Functional Mapping of Quantitative Trait Loci Underlying the Character Process: A Theoretical Framework. *Genetics* **161**, 1751–1762 (2002).

108.   Cappa, E. P. *et al.* Impacts of Population Structure and Analytical Models in Genome-Wide Association Studies of Complex Traits in Forest Trees: A Case Study in Eucalyptus globulus. *PLoS ONE* **8**, e81267 (2013).

109.   Allwright, M. R. *et al.* Biomass traits and candidate genes for bioenergy revealed through association genetics in coppiced European Populus nigra (L.). *Biotechnology for Biofuels* **9**, 195 (2016).

110.   Lamara, M. *et al.* Genetic architecture of wood properties based on association analysis and co-expression networks in white spruce. *New Phytologist* **210**, 240–255 (2016).

111.   Thornton, K. R., Foran, A. J. & Long, A. D. Properties and Modeling of GWAS when Complex Disease Risk Is Due to Non-Complementing, Deleterious Mutations in Genes of Large Effect. *PLoS genetics* **9**, e1003258 (2013).

112.   De La Torre, A. R. *et al.* Genomic architecture of complex traits in loblolly pine. *New Phytologist* **221**, 1789–1801 (2019).

113.   Orr, H. A. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935–949 (1998).

114.   Otto, S. P. & Jones, C. D. Detecting the undetected: estimating the total number of loci underlying a quantitative trait. *Genetics* **156**, 2093–2107 (2000).

115.   Thavamanikumar, S., Southerton, S. G., Bossinger, G. & Thumma, B. R. Dissection of complex traits in forest trees—opportunities for marker-assisted selection. *Tree Genetics & Genomes* **9**, 627–639 (2013).

116.   Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics* **101**, 5–22 (2017).

117.   Müller, B. S. F. *et al.* Independent and Joint-GWAS for growth traits in Eucalyptus by assembling genome-wide data for 3373 individuals across four breeding populations. *New Phytologist* **0**, (2018).

118.   Hall, D., Hallingbäck, H. R. & Wu, H. X. Estimation of number and size of QTL effects in forest tree traits. *Tree Genetics & Genomes* **12**, 110 (2016).

119.   Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).

120.   Bernal Rubio, Y. L. *et al.* Meta-analysis of genome-wide association from genomic prediction models. *Animal Genetics* **47**, 36–48 (2016).

Data Availability

All the latent traits, genotypic data, SNP position files the association mapping scripts used for the analysis are publicly available at are available from zenodo.org at

https://doi.org/10.5281/zenodo.3374415 (DOI 10.5281/zenodo.3374414) . Raw sequence data for all the samples utilized in the study are found through the European Nucleotide Archive under accession number PRJEB29652. The Norway spruce genome assemblies and resources are available from http://congenie.org/pabiesgenome.

Competing interests

The authors have no competing interests as defined by Nature Research, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Author contribution

JB, LZ, SOL, EJM and MRGG (María Rosario García-Gil) all wrote the main manuscript and performed data analysis. TM,NF, TG, LO, BK and HXW collected and analysied the phenotypic data. All authors reviewed the article.

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2020:41

This thesis evaluates the potential of genomic-based breeding in Norway spruce. At this stage, the genetic information rendered by GWAS is insufficient to conduct efficient marker-assisted selection, however it has advanced our knowledge of the genetic architecture of traits of economic and ecological value. On the other hand, GS is considered as a powerful alternative to genomic breeding in Norway spruce.

Linghua Zhou

Department of Statistics

University of Kentucky

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.