2018

Quantitative Genetics of Populus

Roba Bdeir
Michigan Technological University, rbdeir@mtu.edu

Copyright 2018 Roba Bdeir

Recommended Citation
https://digitalcommons.mtu.edu/etdr/709

Follow this and additional works at: https://digitalcommons.mtu.edu/etdr

Part of the Biochemistry Commons, Biotechnology Commons, and the Molecular Biology Commons
QUANTITATIVE GENETICS OF POPULUS

By
Roba Bdeir

A DISSERTATION
Submitted in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
In Biochemistry and Molecular Biology

MICHIGAN TECHNOCLOGICAL UNIVERSITY
2018

© 2018 Roba Bdeir
This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Biochemistry and Molecular Biology.

School of Forest Resources and Environmental Science

Dissertation Co-Advisor: Oliver Gailing

Dissertation Co-Advisor: Victor Busov

Committee Member: Gerald Tuskan

Committee Member: Thomas Werner

School Dean: Andrew J. Storer
Dedication

Dedicated to my beloved parents

Mohammad Aiman Bdeir

&

Randa Alsahli (1974-2009)
Table of Contents

QUANTITATIVE GENETICS OF *POPULUS*................................................................. 1

Dedication .......................................................................................................................... 3

Table of Contents .............................................................................................................. 4

Preface ................................................................................................................................ 7

Acknowledgments ............................................................................................................. 8

Abstract .............................................................................................................................. 9

Chapter 1: General Introduction .................................................................................. 10

Goals and Objectives ..................................................................................................... 13

References ..................................................................................................................... 14

Chapter 2: Quantitative Trait Locus mapping of *Populus* bark features and stem diameter .......................................................................................................................... 16

Abstract ......................................................................................................................... 16

Introduction ................................................................................................................... 17

Materials and methods .................................................................................................. 19

Mapping population .................................................................................................. 19

Plant material and construction of genetic linkage map ........................................... 19

Phenotypic measurements ......................................................................................... 19

QTL analysis .............................................................................................................. 20

Candidate genes ........................................................................................................ 21

Position of LBD genes ............................................................................................... 21

Results ........................................................................................................................... 22

Analysis of phenotypic correlations among traits and trait frequency distributions 22

QTL analysis and detection across contrasting environments ............................... 22

Description of QTL clusters ...................................................................................... 23

Candidate gene identification and characterization .................................................. 27

Discussion ..................................................................................................................... 28
## Chapter 3: Genome-Wide Association Studies of bark texture in *Populus tricocarpa*

### Abstract

47

### Introduction

48

### Material and Methods

- **Plant material**
- **Genetic data**
- **Phenotypic measurements**
- **Statistical Analysis**
- **Genome-wide association mapping**
- **Candidate genes**

50

### Results

- **Phenotyping for bark texture**
- **Genetic associations with phenotypic variation in bark texture**
- **Characterization of genomic regions**

51

### Discussion

- **Radial growth and tissue differentiation**
- **Suberin accumulation**
- **Programmed cell death**

53

### Future Work

58

### Conclusions

59

### References

60

### Figures and Tables

65
Preface

The second chapter is a manuscript published in BMC Plant Biology (Bdeir et al, 2017, DOI 10.1186/s12870-017-1166-4), on which I am first author. My co-advisors, Dr. Oliver Gailing and Dr. Victor Busov, and co-authors Dr. Yordan Yordanov, Dr. Wellington Muchero and Dr. Gerald Tuskan helped conceive and design the study. All provided helpful input on the presented paper and helped in writing and revising the manuscript.

The third chapter comprises a manuscript submitted to BMC Plant Biology on which I am first author. My co-advisors, Dr. Oliver Gailing and Dr. Victor Busov, and co-authors Dr. Yordan Yordanov and Dr. Gerald Tuskan helped conceive and design the study. Dr. Wellington Muchero helped by running the association mapping tests. All provided helpful input on the presented paper and helped in writing and revising the manuscript.

The fourth chapter comprises a manuscript draft to be submitted to Tree Genetics and Genomics on which I am first author. My co-advisors, Dr. Oliver Gailing and Dr. Victor Busov, and co-authors Dr. Matias Kirst, Dr. Wellington Muchero and Dr. Gerald Tuskan helped conceive and design the study. All provided helpful input on the presented paper and helped in writing and revising the manuscript.
Acknowledgments

I would like to extend thanks to the many people, in many countries, who so generously contributed to the work presented in this thesis.

Special mention goes to my enthusiastic supervisors, Oliver Gailing and Victor Busov. My PhD has been an amazing experience and I thank you both wholeheartedly, not only for their tremendous academic support, but also for giving me so many wonderful opportunities. I am also hugely appreciative for sharing their expertise so willingly.

Similar, profound gratitude goes to Gerald Tuskna and Wellington Muchero, who has been a truly dedicated mentor. I am particularly indebted to Wellington for his constant faith in my work, and for his support when so generously hosting me in Oregon. I have very fond memories of my time there.

Special mention goes to Thomas Werner, Matias Krist, Yordan Yordanov and Soha Albukhari, for helping me along my journey.

Finally, but by no means least, thanks go to mom, dad, Moe, my husband Ahmad and my bundle of joy Randa for their unbelievable support. They are the most important people in my world and I dedicate this thesis to them.
Abstract

Evidence for bark, stem and stomatal density adaptation to different climates in the model species *Populus* is seen in both the natural population as well as in the greenhouse, but the genetic basis of these adaptation remains poorly understood. The present thesis investigates bark texture, bark thickness, diameter and stomatal density variations among *Populus* population using two quantitative genetics methods to attempt understand the genetic system controlling inheritance of these traits and to associate them with respective genes. The first approach aimed at detecting quantitative trait loci (QTL) associated with all phenotypic traits in an interspecific hybrid pedigree (*Populus trichocarpa* x *P. deltoides* and *P. deltoides*) collected across several years and sites. Identifying QTL for these important traits will pinpoint to polymorphisms linked to functional genes and provide a list of candidate genes and a baseline for future work. QTL specific to bark texture were highly reproducible in shared intervals across sites, years and replicates. Furthermore, significant positive correlations and co-localization between traits QTL suggest pleiotropic regulators or closely linked genes. Since bark texture showed promising results using the low-resolution QTL method, it was then analyzed further using high-resolution population genomic approach, the genome-wide association study (GWAS). Association mapping defined the genomic regions associated with natural variation in bark texture in a clonally replicated provenance trials of *P. trichocarpa* across three sites, multiple years. The association mapping used with a high-density SNP array allowed us to detect narrow genomic intervals (1-20 kb) with high reproducibility and shared candidate genes for bark texture variation. Several candidate genes were identified related putative function and their location close to QTL maxima were highlighted and are worth further investigation using functional genomics or forward genetics approaches. The results should be exploited for the future conservation and breeding of *Populus* species.
Chapter 1: General Introduction

The genus *Populus*, belonging to the family of Salicaceae, contains 30 to 35 species recognized by taxonomists [1] are forest trees (poplars, cottonwoods, aspens) with tremendous economic and ecological value, as well as valuable characteristics of basic scientific interest. Due to the inherent difficulties of studying such very large long-lived organisms, a model system for forest trees was developed. Poplars have several advantages as a model tree, including fast growth, ease of propagation and of cloning, propensity to hybridize, a small genome and many more [2]. Furthermore, their multiple beneficial uses to society are countless by providing wood, pulp and fiber, contributing to a sustainable development, as well as restoration of forest land and mitigation of climate change.

The spread of fast-growing poplar hybrid plantations across the world was the results of the increasing worldwide demand for wood products as well as the rising world population growth. Breeding programs were established in several countries including Belgium, France, Netherlands, Germany, UK, Canada and the USA [3, 4] to improve poplars for growth, wood products, bioenergy, biomass and environmental enhancement. In China alone, 20 million ha of poplar and willows tree plantings were established by 1978 known as the “Great Green Wall” in efforts to compensate for the many cut forests, resort vegetation and to protect the cities from dust and sand storms [5]. While in the western USA, since the early 1980s private industry planted large blocks of hybrid poplars totaling 20,000 ha [6].

The common cottonwoods of North America (*Populus deltoides* Bartr. ex Marsh and *P. trichocarpa* Torr. & Gray) and the black poplar of Europe (*P. nigra* L.) are the most important species for poplar breeding programs worldwide [7]. *P. deltoides* naturally thrives in the eastern, southern and mid-western USA and southern Canada and has been regarded as the fastest-growing native tree in North America [8], while *P. trichocarpa* occurs in Pacific coastal areas from California to Alaska in the northwestern North America [8]. The hybrid *P. trichocarpa × P. deltoides* shows hybrid vigour (heterosis) with many improved qualities including rapid growth rate, excellent form, site adaptability, ease of propagation and genetic pliability [7, 8]. The knowledge from the pioneering poplar genetic and silvicultural work on *P. trichocarpa × P. deltoides* hybrids
along with the recent release of genomic resources and the *P. trichocarpa* genome [9] will further support the identification of the underlying genetics involved in complex traits important for plants breeding.

In forest genetics, poplars have been adopted by tree physiologists as a model system due to this rapid growth, ease for clonal propagation and strong heterosis in interspecific hybridization [2, 10, 11]. Furthermore, it offers abundant genetic variation in tree morphology, anatomy and physiology where poplars are known to their plasticity and adaptations to varied environments and specific conditions [12]. For these reasons and many more, *Populus* is the center of forest genetics studies from molecular to ecological studies to investigation of large-scale systems to improve our understanding of tree growth and development to use in protection of forest ecosystems and practical tree breeding.

The traditional tree improvement for the selection of trees with imperative traits of economic importance an impractically slow and expensive process due to the long generation interval and since selection is based on physical traits [2, 11, 12] and since most poplars will not flower earlier than 4 years old. Additionally, classical genetic tools such as inbred lines cannot be produced rapidly enough due to the inability of self-pollination. Consequently, we need a sustainable program for this long-lived tree with basic knowledge of its genetic infrastructure including breeding and maintenance of living clone banks. Interspecific hybridization to produce large progeny arrays and multi-generation pedigrees are the basis for genetic mapping experiments designed to identify important genomic regions and the underlying genetics affecting important tree phenotypes [2, 10, 11]. Having the whole-genome sequence of *P. trichocarpa* readily available and the use of genetic markers in the naturally occurring molecular diversity provided an excellent resource for population and quantitative genetics and genomics of the genus.

Molecular markers are specific DNA sequences especially useful in designing breeding strategies to characterize and evaluate genetic diversity in a population, specifically aiding in maximizing genetic gain and monitoring the efficiency of tree improvement traits [13-15]. They are of significant value in multiple trait breeding and are used more efficiently in applied plant breeding, such as breeding for two traits found
in two separate parents are chosen to produce offspring with both desired traits [14]. The
discovery of gene linkage, genes on the same chromosome, led to the establishment of
linkage maps further aided by the development of DNA molecular marker technology
[16, 17]. DNA markers are known to be valuable tools for crop improvements in rice,
wheat, maize, barely and many more plant species [18-21]. As for the genus Populus,
several types of genetic markers and genetic maps have been developed [7, 22, 23].

Genetic mapping, either QTL (Quantitative Trait Loci) or GWAS (Genome Wide
Association Study) is based on the fact that during meiosis genes and markers segregate
via chromosome recombination; thus, genes/markers that are close together (linked) will
be transmitted together from parent to progeny more frequently. This leads to the
determination of the recombinant frequency and converting them into centi Morgans
(cM) as a genetic map unit [24-26]. Once the population is selected, either F2
recombinant inbred lines or a naturally occurring population, the identification of DNA
markers, polymorphisms, between individual trees is critical for constructing a linkage
map. The final step involves marker scoring and coding on individual tree within a
population and conducting linkage analysis using computer programs and creating a
linkage map [24-26]. It is used to analyze and detect an association between the trait of
interest and the genotype of markers pointing to the underlying genes responsible for the
desirable trait.

The conventional QTL mapping is based on the association between genotype and
phenotype in a F2 population, derived by backcrossing the F1 hybrid to one of the parents
[25]. QTL mapping has high statistical power for detecting a QTL, however, it provides
low resolution. Furthermore, only allelic diversity that segregates between the parents of
the particular F2 cross can be assayed; thus, we are limited to the genetic diversity present
in the parents of our segregating population [25]. The second type GWAS or association
mapping (also named linkage disequilibrium mapping) is based on identifying genes
from unorganized natural populations. The low levels of linkage disequilibrium (LD) in
natural populations of Populus makes candidate gene association studies very accurate
since polymorphisms are expected to be within <1 kb of the SNP (single nucleotide
polymorphisms) used to detect it [12]. Association mapping offers a very fine resolution,
to the base pair variation, thus overcoming the limitations found in QTL mapping
technique. Nonetheless, it has its own drawbacks, such as losing power for detection for any rare alleles since the power for detecting a loci using GWAS will be determined by the frequency of alleles [27, 28]. Consequently, GWAS are often complementary to QTL mapping when conducted together they mitigate each other’s limitations.

**Goals and Objectives**

The primary goal of this study was to identify allelic effects of genes controlling economically important traits, including bark and stomatal density, using two population genomics methods, the conventional QTL mapping and association mapping (GWAS). The main objectives of the present study were:

1. Study the genetic variation among different clones and populations of *Populus* across replicates and years.
2. Identify positional candidate genes that underlie QTL for bark texture, bark thickness, diameter growth and stomatal density.
3. Test reproducibility and consistency of QTL across years and environments.
4. Comparison between GWAS mapping with QTL mapping results and quantification of the number of candidate genes within the QTL intervals.
5. Identification of the differences in the genetic control of bark texture in *P. trichocarpa* and *P. deltoides*. 
References


Chapter 2: Quantitative Trait Locus mapping of *Populus* bark features and stem diameter

Abstract

Bark plays important roles in photosynthate transport and storage, along with physical and chemical protection. Bark texture varies extensively among species, from smooth to fissured to deeply furrowed, but its genetic control is unknown. This study sought to determine the main genomic regions associated with natural variation in bark features and stem diameter. Quantitative trait loci (QTL) were mapped using an interspecific pseudo-backcross pedigree (*Populus trichocarpa* x *P. deltoides* and *P. deltoides*) for bark texture, bark thickness and diameter collected across three years, two sites and three biological replicates per site. QTL specific to bark texture were highly reproducible in shared intervals across sites, years and replicates. Significant positive correlations and co-localization between traits QTL suggest pleiotropic regulators or closely linked genes. A list of candidate genes with related putative function, location close to QTL maxima and with the highest expression level in the phloem, xylem and cambium was identified. Candidate genes for bark texture included an ortholog of *Arabidopsis* ANAC104 (PopNAC128), which plays a role in lignified fiber cell and ray development, as well as Pinin and Fasciclin (PopFLA) genes with a role in cell adhesion, cell shape and migration. The results presented in this study provide a basis for future genomic characterization of genes found within the QTL for bark texture, bark thickness and diameter in order to better understand stem and bark development in *Populus* and other woody perennial plants. The QTL mapping approach identified a list of prime candidate genes for further validation using functional genomics or forward genetics approaches.

---

1 The material contained in this chapter was previously published in *BMC Plant Biology* (Bdeir et al, 2017).
Introduction

Bark, the outermost surface of stems and branches in woody plants, encompasses all tissues outside the vascular cambium and includes the secondary phloem, secondary cortex and the periderm [1, 2]. Bark’s outer layer, or phellem, is composed of mostly dead tissues that form a protective barrier between plant and the abiotic and biotic environment, while the inner layer, or phloem, serves as a conduit for transport and storage of photosynthate [1, 2]. Despite its important roles including photosynthate transport [3], photosynthesis [4, 5], storage [6], mechanical support [7] and protection [8-11], the molecular basis of bark formation remains poorly understood (for reviews see: [12, 13]).

Bark texture varies among species, and even among genotypes within species, and has notable phenotypic diversity ranging from smooth, peeling, fractured, fissured to plated [14]. Within genera, bark texture differs between related species, e.g., in mature Populus trichocarpa (Torr. & Gray), bark is smooth or lightly flaky, while in P. deltoides (Bartr. ex Marsh), bark is rough and highly furrowed [15]. Bark’s high morphological diversity suggests that variation in texture may be an important component of variation in plant ecological strategies. It has been reported that in ash and beech smooth bark genotypes are less susceptible to insect and fungal diseases [16, 17]. In addition, bark thickness and moisture content are correlated with enhanced fire resistance [9, 18, 19] and in cork oak the phellem is also the basis of the cork manufacturing industry [20]. Despite the biological, ecological and industrial value of bark, the genetic basis of bark’s features remains undefined.

To further understand the variation in bark texture, we need a better understanding of outer bark development. The outer bark includes all tissues formed by the phellogen, consisting of dead hollow cork cells [1, 2], and originating from the outermost layer of the secondary phloem [2]. Romero [21] has proposed that discontinuous periderms may be the result of variation in radial meristematic activity in the phellogen in apparent response to the mechanical stresses imposed by radial growth, whereas, smooth textured barks may be derived from the formation of a single periderm and continuous shedding of phellem. However, since most plant species develop several periderms over the course of time, smooth bark scales can develop from preceding
periderms from beneath the stem surface while uneven thick and thin layers result in the bark splitting, and in a peeling bark appearance [21]. Finally, Romero [21] also suggests that scaly and fissured bark develops when bark growth is discontinued and overlapping layers of periderms are formed. While these descriptions provide indications on how bark texture can vary, there remains a lack of understanding of phellem development at the molecular level. Additionally, variation in radial meristematic activity in the cambium may affect both diameter growth and bark texture as result of mechanical stresses.

Especially Lateral Organ Boundaries (LBD) genes were found to be important regulators of woody perennial growth in poplar [22]. Specifically, the two LBD genes, PtaLBD1 and PtaLBD4 are expressed at the cambium/phloem boundary and are involved in ray cell and secondary phloem development. Two other LBD genes, PtaLBD15 and PtaLBD18, are expressed at the cambium/xylem boundary and are involved in secondary xylem development.

Quantitative Trait Locus (QTL) mapping in segregating populations is a powerful tool to 1) uncover genes underlying naturally occurring phenotypic variation and 2) dissect the genetic basis of phenotypic traits [23]. QTL-based approaches have often been implemented to study the complex genetic architecture underlying wood formation, including lignin, diameter, height, biomass and various wood chemistry traits [24-31]. However, only a few studies have explored bark features and are limited to bark thickness in *Eucalyptus globulus* [32], *Pinus* hybrids [33] and *Boehmeria nivea* [34]. In *Populus*, QTL mapping and gene expression analyses were used to link sequence polymorphisms and variation in transcript levels [35-37].

Thus, in this study, we investigate bark texture, bark thickness and diameter variation in the pseudo-backcross Family 52-124 derived from a cross between a *P. trichocarpa* x *P. deltoides* hybrid and *P. deltoides* [25, 38]. Novaes et al. [25] performed a QTL mapping study in the same mapping pedigree for 20 biomass and wood chemistry traits (including stem diameter) under different nitrogen treatments and identified a total of 63 QTL distributed across 14 chromosomes. In the present study, QTL mapping was done using a genetic map with high marker density anchored to the *P. trichocarpa* whole-genome assembly [39], which allowed us to determine the map position of QTL and identify underlying candidate genes. Specifically, we performed QTL analyses for bark
texture, bark thickness and stem diameter and report results across three years and two
geographic locations. These three traits have been chosen since we suspect an inter-
connection between their development. Specific objectives were to: 1) identify positional
candidate genes that underlie QTL for bark texture, bark thickness and diameter growth
and 2) test reproducibility and consistency of QTL across years and environments.

Materials and methods

Mapping population

An interspecific hybrid poplar pseudo-backcross pedigree (Family 52-124)
composed of 396 genotypes was created by crossing the hybrid female clone 52-225
(TD), an F1 hybrid derived from P. trichocarpa (TT, clone 93-968) × P. deltoides (DD,
clonc ILL-101), with P. deltoides (DD, clone D124) (Fig. S1a) [25]. The F1 hybrid, clone
52-225, had smooth bark and was crossed with P. deltoides clone D124 with rough bark.
The mapping population was planted at Boardman, OR (45°50′8″N, 119°33′48″W) in
2010 with two replicates of 396 genotypes in a three-block replication for a total of six
ramets per cloned genotype (Fig. S1b). The same pedigree was planted in Morgantown,
WV (39°39′32″N 79°54′19″W), in 2006 with four replicates of the 396 genotypes used in
this study.

Plant material and construction of genetic linkage map

We used the genetic map of the mapping family 52-124 comprised of 3,568 SNP
markers with known genomic positions for QTL identification. SNP genotyping, marker
curation and genetic map construction were previously described by Muchero et al. [31].

Phenotypic measurements

Phenotypic data for bark texture (BT), bark thickness (BTh) and diameter at
breast height (D) for all 396 full-sibs were analyzed in this study. Specifically, for the
Oregon site (OR), bark texture data were collected in year 3, 4 and 5 by visual inspection,
whereas diameter and bark thickness data were collected only in year 3 by using diameter
tape and a bark thickness gauge on two opposites sides of the stem. For the West Virginia
site (WV), bark texture data were collected in year 4 and 6 by visual inspection; diameter
and bark thickness traits were collected only for year 4 using a caliper and ruler. Bark texture was assigned a qualitative score based on a scale from 1 (smooth) to 4 (furrowed with deep grooves) (Fig. 1). Especially, the replicated multi-year measurements allowed for a reliable identification of QTL in Boardman, Oregon. Some QTL in Morgantown, Virginia, could have remained undetected as result of the lower number of replicates and years.

Pair-wise phenotypic correlations were calculated as Pearson correlation coefficients for all three traits across the two different geographic regions and the three years using WinSTAT software [40] to assess covariance within blocks, and within and between years (Table S1). To assess variation within blocks, the two replicates for each of three blocks, named as 1a, 1b, 2a, 2b, 3a & 3b, respectively, were analyzed at the OR site in year 3. In year 4 and 5, only block 1 and 2 were measured.

QTL analysis
The data were evaluated for the presence of outliers and recording errors were corrected or deleted. WinSTAT [40] was used to check for normal distribution of residuals. Transformations were deemed unnecessary (Fig. S2). The traits were analyzed with the interval mapping method implemented in MapQTL6 [41] to map putative and suggestive QTL intervals on the genetic linkage map and to test for reproducibility across years and environments. Composite interval mapping with the Multiple-QTL Model (MQM) method was then used to further refine the QTL regions. Markers closely linked to a putative QTL were selected as cofactors and the selected markers were used as genetic background controls in subsequent MQM mapping. We selected additional cofactors until no additional QTL was detected. Mean phenotypic values across the two replicates per site were analyzed separately for each of the three blocks across three years for the OR site and across two years for the WV site. The putative QTL were subjected to 1,000 genome-wide (GW) and chromosome-wide (CW) permutation tests [42] to determine LOD significance thresholds at the 0.05 significance level (Table S2). A putative QTL was declared when it was detected in at least two replicates or in one replicate in different years or sites, with at least one of those instances exceeding the chromosome-wide LOD threshold. To account for minor deviations from normality in
some cases, we also performed the non-parametric Kruskal-Wallis test which is the equivalent to the one-way analysis of variance [43].

**Candidate genes**

Genes underlying genome-anchored QTL intervals (Table S3, S4) were identified from the *Populus* genome assembly V3.0 [44] in the Phytozome database via BioMart tool (https://phytozome.jgi.doe.gov). A complete gene list with InterPro descriptions was collected including both putative and unknown functions. Expression profiles of the gene models from various tissues: bud, leaf, various parts of root and stem (expression FPKM), were downloaded from the publicly available PhytoMine database (https://phytozome.jgi.doe.gov/phytomine/begin.do). The Affymetrix microarray expression raw date profiles for the traits bark and mature phloem, developing phloem, cambium, developing cambium and mature cambium were obtained from the NCBI, GEO database (https://www.ncbi.nlm.nih.gov/geo/, GEO accession number GSE30507) [45]. The raw data were normalized using the RMA algorithm [46] and further analyzed statistically using TM4:MeV software [47, 48], utilizing Affymetrix probe annotation [49].

The genes’ expression in developing and mature phloem/xylem and cambium was then assessed for each QTL interval for all traits based on publicly available data. For each QTL cluster, genes in the map interval with the highest LOD score and high expression in phloem, cambium and xylem tissues (above the 90th percentile) were compiled in a list (Table S5).

**Position of LBD genes**

The position of the Lateral Organ Boundaries Domain (LBD) genes with putative role in bark development and diameter growth were identified by using the BLAST tool in the *Populus* genome assembly V3.0 in the Phytozome (https://phytozome.jgi.doe.gov/) database against well-established *Arabidopsis thaliana* LBD genes.
Results

Analysis of phenotypic correlations among traits and trait frequency distributions

Bark phenotypes ranged from smooth (1) to deeply furrowed bark (4) (Fig. 1). Shallowly fissured bark typical for *P. trichocarpa* was not found in this backcross pedigree (*P. deltoides* x *P. trichocarpa* hybrid backcrossed with *P. deltoides*). The interspecific crossing parent, *P. trichocarpa* (clone 93-968) x *P. deltoides* (clone ILL-101), had relatively smooth bark (mean value: 1.33, SD: 0.47). The other crossing parent, *P. deltoides* clone D124, had a rough bark texture (mean: 2.66, SD: 0.47). The grandparent *P. deltoides* (DD, clone ILL-101) had a slightly furrowed bark texture while grandparent *P. trichocarpa* (TT, clone 93-968) had smooth and slightly fissured bark (field observations on adult trees, no genotypes and measurements available in the field trials). Bark texture showed the highest correlation within blocks (r=0.91 to 0.93, p≤0.0001). The phenotypic correlations at the OR site within the same year for bark texture ranged from r=0.58 to 0.76; for diameter, r=0.38 to 0.45 and for bark thickness, r=0.40 to 0.56, all at p≤0.0001 (Table S1). Comparing mean values of traits among years, bark texture values were significantly correlated among years at the OR site (r=0.51 to 0.77, p≤0.0001). The correlations were weaker for the WV site, but still highly significant (r=0.39, p≤0.0001). Finally, at year 3, bark thickness showed a strong positive correlation with both bark texture (r=0.32 to 0.69, mean value r=0.49, p≤0.0001) and diameter (r=0.17 to 0.75, mean value r=0.43, p≤0.0001) within the OR site (Table S1, blue and green sections), however, bark texture and diameter showed inconsistency in correlation values and significance ranging from r=0.15 (p≤0.05) to r=0.47 (p≤0.0001) (Table S1, red section). Overall, traits showed high correlations among replicates and years, and traits were correlated with each other. Across site correlations were only significant for bark texture ranging from r=0.25 to 0.40 (p≤0.01-0.0001).

QTL analysis and detection across contrasting environments

Seven major QTL clusters were detected for bark texture on seven individual chromosomes I, II, VI, VIII, XIII and XVIII (Table S2), with all clusters containing at least three individual QTL above the GW threshold. For diameter, three QTL clusters with significance above the GW threshold were detected on chromosome I, VI and
XVIII; in addition, two suggestive QTL above the CW threshold were detected on chromosome VIII and XII (Table S2). Bark thickness showed three QTL clusters above the GW threshold on chromosome I, VI and XVIII, and four QTL above the CW threshold on chromosome II, VIII and XII (Table S2). Chromosome VIII likely contains two separate QTL since they map to distinct chromosomal positions. All 94 individual QTL detected for the three traits across various chromosomes were successfully anchored to the *Populus* genome assembly (Fig. 2; Table S2). For the seven bark texture QTL clusters, the percentage of phenotypic variance explained (PVE) ranged from 3.6 to 12.8% for QTL above the GW threshold, while for diameter and bark thickness, it ranged from 5.4 to 8.4% and 4.5 to 9.6%, respectively (Table S2). For QTL on chromosome II, VI, VIII and XII the *deltoides* genotype DD was associated with the lower value for bark texture (Table S2), while for QTL on chromosome I, XIII and XVIII the DT genotype was associated with a lower value for bark texture.

Figure 3 shows a graphical outline of LOD score profiles for bark texture QTL versus map location across all seven chromosomes before (left) and after cofactor selection (right). The QTL for bark thickness and diameter, described above, overlap with six out of the seven bark texture QTL clusters using the interval mapping approach (Fig. 2). Specifically, QTL for bark texture overlap with diameter and bark thickness QTL on chromosome I, VI and XII, and solely with bark thickness QTL on chromosome II, VIII and XVII (Fig. 2). Overall, reproducibility and co-location within the same map interval are observed across experimental replicates within sites and years and in some cases across sites.

Description of QTL clusters

Based on significance, consistency and reproducibility of the QTL across sites, years and replicates, specifically for bark texture, we classified the QTL clusters according to four criteria: 1) significance (LOD scores), 2) reproducibility across biological replicates, 3) reproducibility over time (years) and 4) reproducibility across environments (sites). All seven QTL clusters were significant with at least three individual QTL having LOD scores above the GW threshold and were reproducible across blocks (biological replicates) within the same year and across two years for the OR
site. Four QTL clusters on chromosome I, II, VIII and XIII, were reproducible between sites across very different environments. In each QTL cluster most of the QTL were associated with bark texture. Below we provide a detailed description of all seven bark texture QTL clusters ranked according to the four criteria and their association with bark texture, bark thickness and diameter.

QTL cluster on chromosome I was associated with all three traits (Fig. 2). For diameter four out the five QTL were significant at the GW threshold and for bark thickness three out of the four QTL were above the GW threshold (Table S2). For bark texture, twelve QTL were detected across all replicates and years in both OR and WV and ranged from 101 to 192 cM (before cofactor selection) with highly reproducible LOD maxima after cofactor selection consistently around 165 cM (Table S2; Fig. 3a). For the OR site, seven out of the ten QTL in this cluster were above the GW threshold all being detected in years 3 and 5; whereas at the WV site, the two QTL for both years were above the CW threshold. All twelve individual QTL mapped reproducibly to the same map interval and LOD maxima positions at the GW threshold were typically associated with no more than three markers in close proximity, around 3 cM (Fig. 3a). Notably, the results showed consistency and high reproducibility, first between blocks within individual years, second within the same site, and third across sites. Overlapping QTL intervals and maxima with high significance across all years are presented in Table S2 and in Figures 2, 3.

The second bark texture QTL cluster was detected across all replicates and years in OR, and in year 4 in WV, and was mapped on chromosome VIII within the chromosomal region 48 to 104 cM before cofactor selection with varied LOD maxima positions after cofactor selection (~50 cM, ~65 cM or 74 cM) (Table S2; Fig. 3a). When comparing this cluster with other traits, one overlapping QTL for bark thickness (significant at the CW level), peaking at 67 cM was detected. For the OR site, six of the ten QTL for bark texture in this cluster were significant above the GW threshold; whereas at the WV site, only one QTL was found above the CW threshold. All eleven individual QTL within this cluster mapped reproducibly within the same map interval, however positions of QTL maxima varied for the OR site (Fig. 3a). The overall reproducibility
within the OR site, across both sites and across all years, was high and QTL were found within the same chromosomal region covering 25 cM.

The third bark texture QTL cluster mapped on chromosome XIII and was detected across all years and replicates in OR, and in year 4 in WV, within the chromosomal region 83 to 118 cM before cofactor selection and reproducible LOD maxima after cofactor selection was found around 108 cM (Table S2; Fig. 3a), which interestingly co-located with the Lateral Organ Boundaries Domain gene LBD15c (Potri.013G156200), a candidate gene for xylem development [22]. No QTL for bark thickness or diameter were found in this QTL interval. For the OR site, seven out of the nine QTL in this cluster were significant above the GW threshold across all years; whereas at the WV site, only one QTL was detected and found above the GW threshold. All ten individual QTL mapped reproducibly within the same map interval and positions of QTL maxima were typically associated with no more than four markers in close proximity covering a map interval around 5 cM (Fig. 3a). The results showed overall consistency and reproducibility between block replicates within individual years, within the OR site and across the two sites and across most years (Table S2, Fig. 3a).

The fourth bark texture QTL cluster on chromosome II was detected in both OR and WV, across all years and replicates for both sites, within a chromosomal interval from 0 to 70 cM (before cofactor selection) but displayed inconsistent and variable LOD maxima after cofactor selection (Table S2; Fig. 3b). When comparing this cluster with bark thickness and diameter, one overlapping QTL for bark thickness (significant at the CW threshold), peaking at 0 cM, was found. Eight out of the ten individual bark texture QTL were significant above the GW threshold, indicating reproducibility within replicates and sites, between sites and across years. However, after cofactor selection, the LOD maxima greatly varied even within replicates of the same year.

The fifth bark texture QTL cluster was detected on chromosome VI in OR in years 3 and 5 only, within a chromosomal interval from 135-204 cM (before cofactor selection) and had LOD maxima after cofactor selection between 153-162 cM (Table S2; Fig. 3b). Several overlapping QTL, four for diameter (two GW QTL) and two GW QTL for bark thickness, were found within the same chromosomal interval, however the LOD maxima varied among traits after cofactor selection. Six out of the seven individual QTL
for bark texture were above the GW threshold, indicating reproducibility within replicates at the OR site and across the two years. However, the results were not consistent across the two sites, as no QTL was detected at the WV site. Only three out of the nine individual QTL were found significant above the GW threshold, while others were significant at the CW threshold.

The sixth bark texture QTL cluster mapped on chromosome XII and was detected across all years and replicates in OR within the chromosomal interval from 62-113 cM before cofactor selection and consistent LOD maxima between 98-104 cM were detected after cofactor selection at the OR site (Table S2; Fig. 3b). Two individual QTL, one for bark thickness (CW QTL) and one for diameter (CW), were partially overlapping with bark texture QTL, but had LOD maxima separate from the LOD maxima for bark texture. Out of the nine individual QTL, only three were above the GW threshold. The results are reproducible between replicates and within and across the three years in OR. Though no significant QTL were detected in WV, the QTL graph shows a suggestive QTL with increasing LOD score near the same chromosomal interval noted above (Fig. 3b), though still below the CW threshold.

Finally, the seventh bark texture QTL cluster, mapped on chromosome XVIII within the chromosomal interval from 13-35 cM, before cofactor selection, with reproducible LOD maxima at 30 cM after cofactor selection, detected in year 3 and 5 at the OR site (Table S2 and Fig. 3b). A QTL cluster for bark thickness closely overlapped with the same chromosomal interval with LOD maxima further upstream. Four separate QTL for diameter were detected on the same chromosome (two above the GW threshold). In year 5, three out of the seven bark texture QTL were above the GW threshold. The QTL were only reproducible across year 3 and 5 at the OR site, although a suggestive QTL was found for WV (Fig. 3b).

The Kruskal-Wallis rank sum test was subsequently used to confirm significant associations of individual markers linked to the QTL. For all traits, the markers underlying the QTL interval were also significantly associated with the traits (p<0.005). The Kruskal-Wallis test provides further confirmation of the marker-trait association, indicating that the results of the QTL analysis were not influenced by segregation distortion or non-normal distribution of certain traits.
Candidate gene identification and characterization

To narrow the QTL position and identify candidate genes within the QTL interval MQM mapping was used. Intervals spanning the genomic regions (physical location by MQM mapping) summarized in Table S2 were used to identify all genes occurring within the seven QTL clusters for bark texture and for the QTL clusters that were associated with bark texture, diameter and bark thickness. The number of genes for each trait in QTL clusters based on MQM mapping with cofactor selection is summarized in Table S3. There were 1869 genes within genome-anchored QTL intervals for bark texture, out of which, 1476 (82%) had annotations based on the InterPro domain and expression profiles (average FPKM) for 22 different tissues and based on Affymetrix microarray expression data for bark and woody tissues (phloem, cambium and xylem), whereas a total of 693 and 789 genes were detected in QTL clusters for diameter and bark thickness, respectively (Table S4).

QTL for all three traits overlap in QTL cluster on chromosome I, where the diameter QTL included 25 genes and the bark thickness QTL encompassed 115 genes. Two additional QTL clusters on chromosome VI and XII were associated with all three traits where diameter and bark thickness QTL had 369 and 209 overlapping genes on chromosome VI and 29 and 14 overlapping genes on chromosome XII, respectively. QTL clusters for bark texture on chromosome II, VIII and XVIII overlapped with bark thickness QTL only, containing 963, 23 and 38 overlapping genes, respectively (Fig. 2, Table S4). As a result of the Salicoid duplication event in the *Populus* genome, nearly every chromosome has a paralogous segment elsewhere in the genome; this is due to the whole-genome duplication between chromosomes resulting in homologous genomic blocks [50]. Each of the seven chromosomes noted above has a Salicoid paralog, yet none of these duplicated genes in paralogous segments co-located with other QTL for the same trait.

Top expressional candidate genes (above the 90\textsuperscript{th} percentile) for mature and developing phloem, cambium, developing xylem and mature xylem in QTL intervals with the highest LOD scores resulted in a compiled list of the top candidate genes for each trait (Table S5). In total, the top candidate genes with putative function in the
control of bark texture, diameter and bark thickness are narrowed down to 40, 20 and 46 genes spanning various QTL clusters (Table S5).

**Discussion**

We have characterized segregating bark features in an interspecific backcross of *Populus*. With the use of QTL mapping, we are able to link the phenotypic traits to their associated polymorphisms in the genome, thus integrating phenotypic and genotypic data to identify putative genetic mechanisms related to phellem development.

While other studies have identified QTL in interspecific *P. trichocarpa* and *P. deltoides* families for many different traits, including leaf size and shape, growth and bud set, diameter, height, stem and root biomass and various wood chemistry phenotypes [25, 35, 51-54], little research has been done on bark features [55-57], despite bark being one of the key-energy-related characteristics of lignocellulosic feedstock [58-60].

We identified several QTL that encompassed both bark traits and stem diameter and found that these intervals mapped consistently across geographic locations, replicates within sites and across years. Interestingly, one study, using the same pedigree, reported several overlapping QTL with our traits [25]. Specifically, bark texture QTL in our study overlapped on chromosome I, II and XVIII with QTL associated with total biomass, C5 and C6 sugars, and height. Additionally, diameter and bark thickness QTL overlapped on chromosome VIII, XII and XVIII with QTL for diameter and biomass traits [25]. It is difficult to determine if there are genes that have pleiotropic effects or whether there are alternate genes within the co-located intervals because of the large size of the interval and lack of expression evidence in the Novaes *et al.* [25] study.

While a few studies have analyzed bark thickness, e.g., *Eucalyptus globulus*, *Pinus* hybrids and *Boehmeria nivea* [32-34], the genetic basis and causal loci of bark thickness and/or bark texture have not yet been determined. In *Boehmeria nivea*, a perennial herbaceous plant belonging to the Urticaceae Family, several QTL for bark thickness have been mapped and some were identified in the same QTL intervals across two contrasting environments in Changsha, China at varying time throughout the year [34].
Bdeir et al. (2016) and Yordanov et al. (2010) previously identified genes with a role in bark development [61, 22]. Based on the generation of loss-of-function phenotypes through transgenic plants, Lateral Organ Boundaries Domain (LBD) genes were found to have a crucial role in meristem maintenance and were identified as important regulators of woody perennial growth in poplar, specifically in *Populus tremula* x *Populus alba* (Pta). The overexpression of PtaLBD1 resulted in wide multiseriate rays as compared to uniseriate rays in the wild type [22]. In the regulation of secondary (woody) growth, two genes of the LBD Family (PtaLBD1 and PtaLBD4) were involved in secondary phloem and ray cell development and two genes (PtaLBD15 and PtaLBD18) in secondary xylem formation. Interestingly, one of the PtLBD15 paralogs (Potri.013G156200, total expression 36.30), previously found to have a role in secondary growth, was found in the QTL cluster on chromosome XIII for bark texture with the highest LOD score. In this chromosome region seven out of a total of 10 QTL detected were above the genome-wide threshold (Fig. 2). PtaLBD1 (Potri.008G043900) was detected in the diameter QTL cluster on chromosome VIII with a LOD score of 2.27. Both PtaLBD15 and PtaLBD1 are involved in secondary growth in poplar and are potential candidate genes for diameter growth and bark characteristics. PtaLBD15 was found in *Populus tremula* x *P. alba* to be mainly expressed at the cambium/xylem boundary and thus is likely involved in secondary xylem development. PtaLBD1 was found to regulate secondary phloem and ray development and was highly expressed in the phloem and cambial zone [22]. Therefore, their apparent involvement in secondary growth and development in poplar and their detection in QTL clusters for diameter and bark texture make them candidate genes for these traits. Finally, PtaLBD12 (Potri.008G072800, total expression 26.06) was detected in the QTL cluster on chromosome VIII and overlapped with diameter and bark thickness QTL. PtaLBD12 has been reported to be involved in the development of various lateral organs from the meristem in *Arabidopsis* plants, but its role in secondary growth is unknown [62].

We were able to identify a list of candidate genes underlying the QTL intervals of all three traits using genetic markers anchored to the *Populus trichocarpa* genome. Generally, QTL were highly reproducible among biological replicates and years and even across geographic locations. While some QTL studies obtained good reproducibility
across two different time data set [33, 34, 63], our study has identified significant QTL consistently co-locating across sites, years and replicates especially for bark texture. For instance, QTL in clusters on chromosome I, VIII, XII, XIII and XVIII were consistently identified within ~25 cM.

When comparing all the aspects of the seven bark texture QTL clusters, including significance, reproducibility across replicates, years and sites, along with consistency of the QTL, major QTL were identified on chromosome I, VIII and XIII, where QTL maxima were found within a 5-20 cM interval across most replicates, years and at both sites. Given the environmental contrast between the OR and WV experimental sites, four out of the seven QTL clusters, representing a total of 47 individual QTL, detected for bark texture were remarkably consistent across both sites. Differences in reproducibility for QTL clusters across sites suggest differential environmental effects on gene expression. In comparison to bark texture, QTL clusters for bark thickness and diameter had lower reproducibility across sites.

In QTL clusters on chromosome I, VI and XII, QTL for bark texture, bark thickness and diameter were syntenic. Co-location of QTL for traits can be the result of pleiotropic effects or closely linked genes. These overlapping QTL could be an explanation of different aspects of bark texture and radial growth. Romero [21] proposed that rough bark results in response to the mechanical stresses imposed by a varied radial growth and due to different meristematic activity in the phellogen, a discontinuous periderm. Strong correlation between bark texture and diameter could indicate that bark texture is partly related to diameter growth. Using interval mapping without cofactor selection, QTL for bark texture overlap with QTL for both traits which are related to radial growth (diameter and bark thickness) on three chromosomes (I, VI, XII) and solely with bark thickness on three other chromosomes (II, VIII, XVII). Furthermore, using MQM mapping, QTL for these traits were mapped to different neighboring positions of the same chromosomes (Fig. 2). Consequently, bark texture seems to be only partly related to diameter growth, and other factors such as meristematic activity of the phellogen and cell adhesion are likely to have major effects on bark texture. A higher mapping resolution as obtained in linkage disequilibrium mapping in natural population
samples is needed to narrow down the QTL region to individual genes and to distinguish between pleiotropic effects and close linkage.

Bark features in our study ranged from smooth to deeply furrowed which is characteristic for *P. deltoides*. Variation in shallowly fissured bark which is characteristic for *P. trichocarpa* was not observed among the segregating progeny. Thus, the QTL identified in this progeny set only represent a subset of a larger number of polymorphisms affecting the traits. And in a pseudo-backcross involving multiple *P. deltoides* parents, polymorphisms associated with characteristic bark features of *P. trichocarpa* seem to be largely undetected. Association populations for *P. trichocarpa* will be used to find additional candidate genes associated with bark texture in this species.

Each of the seven QTL clusters detected have a Salicoid paralog, and yet none of these paralogous genes showed up in the QTL analyses as significant. This further supports that the identified chromosomal regions are not artifacts of spurious correlations. Due to large genomic intervals in QTL clusters with partly overlapping QTL intervals the identification of specific candidate genes was difficult. This limitation was evident in our analyses where only two out of the seven QTL clusters on chromosome XIII and XVIII, encompassed less than 70 candidate genes, while the other clusters included from 123 to 963 candidate genes. Nonetheless, several candidate genes within the QTL interval can be identified based on their putative functions. Using the MQM method, we were able to identify informative loci for bark texture and narrow the QTL region to a small chromosomal region with a short and manageable candidate gene list. For example, a total of 11 NAC genes were detected in bark texture QTL clusters one, five, one, three and one paralogs found on chromosome I, II, VI, VIII and XII, respectively. Specifically, the gene PopNAC128 (Potri.001G206900) is a prime candidate gene and was identified in the QTL interval on LG I within the QTL maxima (LOD score 9.88) and with a moderate expression value. PopNAC128 is one of the orthologs of *Arabidopsis* ANAC104 (*Arabidopsis* Nac Domain Containing Protein 104) and XND1 (Xylem NAC Domain 1). In a related study, *Populus* and *Arabidopsis* transgenic plants with overexpression of these genes resulted in severe dwarfing, lacking phloem fibers and a reduction in stem diameter, cell size and number, vessel number, and frequency of rays in
the xylem [64]. While this study did not focus on bark texture, lack of sufficient lignified fiber cells in the mutant affects the development of fiber bundles and ultimately bark texture as result of slowed secondary phloem development.

Another interesting gene, Potri.001G206700, an ortholog of AT4G33430 (BAK1, Bri1-Associated Receptor Kinase; ELONGATED; SERK3), is involved in patterning and growth regulation [65-67] and was found in the QTL interval on chromosome I, also within the QTL maxima (LOD score 9.88) and with a very high expression value above the 90th expression percentile across phloem/xylem and cambium tissues.

Variation in bark texture could be related to cell adhesion, which is essential to form a single periderm resulting in smooth bark, while lack of cell-cell adhesion leads to the development of uneven and discontinued bark or bark splitting causes a peeling and fissured bark appearance [21]. At the molecular level, several QTL and expressional candidate genes with high expression in phloem/ xylem and cambium identified in this study have a role in cell adhesion, including Pinin (Potri.001G208200) and PopFLA or Fasciclin-Like Arabinogalactan (Potri.013G151300, Potri.013G151400 and Potri.013G151500). Interestingly, both genes fall within the QTL interval with the highest LOD scores and are above the 90th gene expression percentile for both xylem and phloem tissues. Many of the studies on Pinin, mainly on animal epithelial cells, revealed a vital role in cell-cell adhesion and cell shape [68, 69]. No studies exploring the Pinin gene in plants were found. The FLA gene is better studied across the plant kingdom, including Arabidopsis and Populus, and shows specific and high expression during the onset of secondary-wall cellulose synthesis, particularly in stem cells undergoing secondary-wall deposition [70, 71]. Transgenic lines indicate a role in cell-wall architecture and composition. Specifically for PopFLA, a role in tension wood formation in the xylem of mature stems was suggested based on a reduction in transcript levels leading to reduced stem flexural strength by modulation of cellulose and lignin composition in the xylem [70, 72, 73].

**Future work**

The top QTL and expressional genes reveal additional potential candidates (Table S5), some of which are proteins of unknown or putative function and have never been
studied. These genes represent potential candidate genes for future studies using either functional genomics or forward genetics techniques. Candidate genes within QTL intervals were identified based on Affymetrix Microarray expression profiles obtained from public databases. In the future, qRT-PCR confirmation of candidate genes’ expression profiles should be performed in various tissues (phloem, cambium, xylem, phellogen) of the parental clones and part of the mapping pedigree.

**Conclusion**

In conclusion, the results presented in this study provide a basis for future genomic characterization of genes found within the QTL for bark texture, bark thickness and diameter in order to better understand stem and bark development in *Populus* and other woody perennial plants. Additionally, profiling the expression of the candidate genes (eQTL studies) in the developing bark of the mapping pedigree would allow parsing the list of candidate genes into those genes with high expression profiles in the tissue of interest. Bark texture is a complex trait which can be affected by differences in cell adhesion and radial meristematic growth. In the future, developmental differences between bark texture phenotypes should be analyzed in anatomical sections in representative genotypes and developmental stages.
References


Figures and Tables

**Figure 1** Bark texture scale for *Populus* Family 52-124 offspring. Ranging from smooth (1), medium (2), rough (3) and rough/deeply furrowed bark (4), and level 4 was only found in Oregon.
Figure 2 QTL anchored to the genome of *Populus trichocarpa* (V3 assembly, Kelleher et al., 2007). The actual map has a high marker density (average marker spacing: 5 markers per 4 cM). For illustration purposes, for each linkage group an evenly spaced selection of scaffolds is shown (1 marker per 20 cM). The yellow regions on LGs represent LOD score maxima across years and environments. QTL for bark texture (BT), diameter (D), and bark thickness (BTh) are shown in red, green, and blue for Oregon and in pink, light green, and turquois for West Virginia and named according to Table S2. LOD score maxima, genome-wide intervals (solid bars) and chromosome-wide intervals are shown for QTL that were identified in different years and environments (see Table S2). The outer lines of bars are CW thresholds and middle lines are QTL LOD maxima. The exact map and physical locations of QTL are shown in Table S2. Scaffold intervals are represented in Mb. Black vertical lines represent the physical location of LBD genes in the *P. trichocarpa* genome, orthologues are notes by a, b or c.
Figure 3 LOD score profiles for bark texture (BT) QTL detected in the *Populus* Family 52-124. Specifically, BT QTL detected on a) LG I, VIII and XIII and b) LG II, VI, XII and XVIII using interval mapping (left) and the Multiple-QTL Model (MQM) with co-factor selection (right) across all experimental replicates at Oregon (OR) across 3 years and at West Virginia (WV) across 2 years. Chromosome-wide (CW) and genome-wide (GW) significance thresholds are shown with dashed lines ($\alpha = 0.05$, 1,000 permutations). Yellow, orange, and red solid lines represent LOD score profiles for 3-year-old replicated samples at Oregon, solid shades of green lines for 4-year-old samples, and shades of blue lines for 5-year-old samples. Dotted red and green lines represent LOD profiles for 4-year-old and 6-year-old samples at the West Virginia site. Two broken horizontal lines represent GW and CW LOD significance threshold after 1,000 permutations at the $p \leq 0.05$ significance level. The exact map and physical locations of QTL are shown in Table S2.
Supplementary Figure 1 Mapping pedigree and block layout for *Populus* Family 52-124 used in this study. a) Family 52-124 is a pseudo-backcross pedigree between clone 52-225, an F1 hybrid derived from *P. trichocarpa* (T), 93–968 X *P. deltoides* (D) (ILL-101), back-crossed to *P. deltoides* (clone D124 (Novaes et al. 2009). Shading in pedigree: black = TT, gray = TD, White = DD genotype. b) Progeny plantation replicates and block layout in Oregon. Two adjacent replicates in each block and each genotype is represented in three blocks (6 replicates in total). Only two genotypes out of 392 genotypes are shown as an example.
Supplementary Figure 2 Frequency distribution for bark texture, diameter and bark thickness (a, b, and c, respectively) across Oregon and West Virginia sites and various years in *Populus* Family 52-124.
Supplementary Table 1² Pair-wise estimates of phenotypic correlations calculated as Pearson correlation coefficients between bark texture (BT), diameter (D) and bark thickness (BTh) phenotypes collected from *Populus* Family 52-124.

Note: Specifically, the analysis was done within and between experimental blocks, across years and across sites. Symbols show p-value significance, where p≤0.05=*, p≤0.01=**, p≤0.001=***, and p≤0.0001=****. White cells are correlations between same trait across years and sites; red cells are correlation between diameter and bark texture; green cells are correlation between bark thickness and bark texture; blue cells are correlation between bark thickness and diameter.

Supplementary Table 2² QTL associated with bark texture, diameter and bark thickness identified in *Populus* Family 52-124 in Oregon and West Virginia.

Note: Chr: chromosome; V2: markers anchored on version 2 of the *P. trichocarpa* genome; V3: version 3 updated physical location; PVE: percent phenotypic variance explained; DD: homozygous for the *P. deltoides* allele, DT: heterozygous for the *P. deltoides* and *P. trichocarpa* alleles. LOD max determined using MQM mapping, value with *: above GW threshold, otherwise above CW threshold. Alternating white or grey shades represent unique QTL found in one or multiple replicates across years and sites. Indexes 1, 2, 3 or m designate: replicate one, two, three or replicates’ mean value for OR samples, whereas for WV samples only one replicate was available; BT: bark texture; D: diameter; BTh: bark thickness; 3y, 4y, 5y or 6y: # years old; OR: Oregon, WV: West Virginia. (Example for population: 1BT3yOR: replicate one, bark texture, 3-year-old samples, Oregon site).

² Supplementary table 1 and 2 are in excel format submitted as digital files.
**Supplementary Table 3** Number of candidate genes detected across QTL for all traits.

<table>
<thead>
<tr>
<th>Trait</th>
<th>QTL on Chromosome</th>
<th>Number of genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bark texture</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>VIII</td>
<td>321</td>
</tr>
<tr>
<td></td>
<td>XIII</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>978</td>
</tr>
<tr>
<td></td>
<td>VI</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>XII</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td>XVIII</td>
<td>47</td>
</tr>
<tr>
<td>Diameter</td>
<td>I (~16820000bp)</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>I (~31760000bp)</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>VI</td>
<td>362</td>
</tr>
<tr>
<td></td>
<td>VIII</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>XII</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>XVIII</td>
<td>231</td>
</tr>
<tr>
<td>Bark thickness</td>
<td>I</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>303</td>
</tr>
<tr>
<td></td>
<td>VI</td>
<td>208</td>
</tr>
<tr>
<td></td>
<td>VIII (~157000bp)</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>VIII (~550000bp)</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>XII</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>XVIII</td>
<td>52</td>
</tr>
</tbody>
</table>

The number of genes for each trait in QTL clusters based on MQM mapping with cofactor selection, sorted by significance and reproducibility.

**Supplementary Table 4** All candidate genes within the ninety-four QTL detected in *Populus* Family 52-124. Physical localization, annotation and expression profile of gene models within each QTL interval for all traits.

**Supplementary Table 5** The 90th percentile candidate genes within the ninety-four QTL detected in *Populus* Family 52-124. Physical localization, annotation and expression profile of gene models in the 90th percentile with high expression within LOD peaks for each QTL interval for all traits.

---

3 Supplementary table 4 and 5 are in excel format submitted as digital files.
Chapter 3: Genome-Wide Association Studies of bark texture in Populus trichocarpa

Abstract
Plants have developed bark as a defense barrier to deal with environmental stresses such as pathogen invasion, drought, fire and UV radiation. Bark is formed when radial expansion pushes the cortex and epidermis outward creating secondary meristem. Mature phellem development requires cell growth, differentiation and cell death as the bark’s outer layer is composed of dead differentiated cells. However, the genetic control of this complex phenotype is generally unknown. Here we use association mapping to define the genomic regions associated with natural variation in bark texture in Populus trichocarpa. Clonally replicated provenance trials of P. trichocarpa were studied for bark texture collected across three sites, multiple years and (two to three) biological replicates per site. A total of 41 genomic intervals containing SNPs significantly associated with bark texture were detected that were highly reproducible across sites, years and replicates. A list of candidate genes within these regions with related putative function was identified. A total of 98 genes were considered candidate genes due to significance and putative function with connection to the bark texture phenotype. Genome-wide association mapping with a high-density SNP array allowed us to detect narrow genomic intervals (1-20 kb) with high reproducibility and shared candidate genes for bark texture variation. For example, a membrane-associated apoptosis protein and a wall-associated receptor kinase (PR5K-like) protein, which both are involved in radial growth and tissue differentiation, lie within highly significant trait-associated regions. Two copies of root hair defective 3 genes, known as transmembrane proteins in Populus differentiating xylem and phloem, were also significantly associated with bark texture and co-located with major QTL for bark texture, bark thickness and diameter from our previous study, suggesting an important role in radial growth.

4 The material contained in this chapter has been submitted to Tree Genetics and Genomics.
Introduction

Bark in trees is formed when radial growth pushes the cortex and epidermis outward creating secondary meristem, i.e., the cork cambium [1]. The cork cambium then outwardly produces cork (phellem) cells, the bark’s outer layer which is composed of dead differentiated cells [1]. As an adaptation to terrestrial life, bark acts as a defense barrier, including suberins in the phellem layers, to deal with environmental stresses such as pathogen invasion, drought, fire and UV radiation [2, 3]. Despite of its importance, tree bark biology has been less extensively studied than wood biology.

Mature phellem development requires cell growth, differentiation and death. There are common steps in the differentiation of lignified secondary xylem, phloem and phellem cells, including cell division, cell expansion or elongation, secondary cell wall deposition, lignification, and programmed cell death [4]. Carpita et al. [5] estimate that as many as 15% of the genes in a plant genome may play a role in cell wall synthesis, remodeling, or turnover. Secondary cell wall formation is a complex process that requires the coordinate regulation of diverse metabolic pathways [6], however the molecular mechanisms and components that may be essential for cell wall formation in phellem are still unclear. Furthermore, while some defects or variations in secondary cell walls have been characterized, such as irregular xylem [7] or defects in lignin biosynthesis [8], the genetic basis of the phenotypic diversity in bark texture remains completely undefined.

Bark texture varies among species, and even among genotypes within species, and has notable phenotypic diversity ranging from smooth, peeling, fractured, fissured to plated [9]. Differences in bark texture are of ecological importance and have been related to different levels of susceptibility to fungal and insect diseases and to fire [10-13]. For example, tree species with smooth bark, e.g., beech and ash, have been reported to be less susceptible to insect and fungal diseases [12, 13]. Bark thickness and moisture content are reported to be associated with higher fire resistance, where a thicker bark allows for their survival during intense fires [11, 14, 15]. Previously, we have analyzed the bark texture of an interspecific pseudo-backcross pedigree derived from a cross of a female hybrid clone (Populus trichocarpa x P. deltoides) with a male P. deltoides parent through QTL mapping [16]. Highly reproducible QTL were detected encompassing several candidate genes including PopNAC128, which has been reported to play a role in
lignified fiber cell and ray development [17], along with Pinin and Fasciclin (PopFLA) that play a role in cell adhesion, cell shape and migration [18, 19].

The parental species used to derive the pedigree display contrasting bark textures. *P. trichocarpa* (Torr. & Gray) bark is smooth or lightly flaky, while *P. deltoides* (Bartr. ex Marsh) bark is rough and deeply furrowed. Because the backcross was with a *P. deltoides* parent, the QTL pedigree segregated mainly for bark types typical for *P. deltoides*. It is therefore unclear, if and which of the discovered QTL are common or specific for the bark types of the two species.

Here, we took advantage of a large *P. trichocarpa* association population established at three field sites to further investigate the genetic architecture of bark texture through a genome-wide association study (GWAS). We compare the QTL intervals with the location of GWAS candidate genes to identify common genes and at the same time study new candidate genes and loci, not discovered in the QTL study and possibly specific to the development of *P. trichocarpa* bark texture. Our specific objectives are as follows: 1) identification of candidate genes for bark texture, 2) comparison to QTL mapping results and quantification of the number of candidate genes within the QTL intervals, and finally 3) identification of the differences in the genetic control of bark texture in *P. trichocarpa* and *P. deltoides*.

**Material and Methods**

*Plant material*

Clonally replicated populations of 1,100 black cottonwood genotypes (*Populus trichocarpa* Torr. & Gray) were collected from the whole species range encompassing ca. 16 native stands stretching from 38.8° to 54.3° N latitude, from California to British Columbia [20] and were established in replicated field trials in Corvallis, OR (44°34′14.81″N 123°16′33.59″W), Clatskanie, OR (46°6′11″N, 123°12′13″W) and Placerville, CA (38°45′14″N, 120°44′25″W) as previously described [21].

*Genetic data*

A total of 917 genotypes from the population were resequenced to a minimum 15x depth using the Illumina Genome Analyzer, HiSeq 2000 and HiSeq 2500 as previously described for the same population [20]. Briefly, reads were aligned to the *P.
trichocarpa reference genome version 3 using BWA 0.5.9-r16 with default parameters and SNPs and indels were called using SAMtools mpileup (-E –C 50 –DS –m 2 –F 0.000911 –d 50000) and bcftools (-bcgv –p 0.999089). The resulting SNP and indel dataset is available at http://bioenergycenter.org/besc/gwas/. From this output, we created a GWAS mapping panel of 8.253,066 million SNPs with a minor allele frequency of 0.05.

Phenotypic measurements

Phenotypic data for bark texture (BT) were analyzed in this study for GWAS in all three sites. For the Clatskanie site (CLA), bark texture data were collected in 2012, 2013 and 2015 [i.e., 3-, 4-, and 6-year-old trees] by visual inspection, whereas for the Corvallis (COR) and Placerville sites (PLC), the data were collected in 2013 and 2014 [i.e., 4- and 5-year-old trees], respectively. Bark texture was assigned a qualitative score based on a scale from 1 (smooth) to 3 (flaky with grooves) (Fig. S1). Two replicates of 1,100 genotypes in three blocks, for a total of six ramets per genotype, were analyzed at CLA13 and PLC14 sites, along with three replicates for CLA12 and COR13 and one replicate for CLA15. In addition to the field assessment, the Clatskanie site collection during 2012 had two replicate wood disks with bark collected at breast height per genotype, thus a total of 5 replicates. Differences in age sampling and across different environments were used to evaluate variation in measured phenotypes.

Statistical Analysis

Pair-wise phenotypic correlations were calculated as Spearman’s rank correlation coefficients for all data sets across the three different geographic regions using WinSTAT software [22] to assess correlations within blocks and within and between years (Table S1).

Genome-wide association mapping

The data were evaluated for the presence of statistical outliers, and recording errors were corrected or deleted. WinSTAT was used to check for normal distribution of residuals. Transformations were deemed unnecessary (Fig. S2). Association analyses
were performed using the efficient mixed model association (EMMA) algorithm in the EMMAX software with kinship as the correction factor for genetic background effects [23] to compute genotype to phenotype associations using 8.2 million SNP and indel variants with bark texture as the phenotypic variable. On average, a marker is represented every 17 base pairs (bp) across the genome. Significant and reproducible SNPs strongly associated with bark texture trait were detected at a FDR threshold of 0.05. Significance thresholds were set at negative log p-values of 6. False discovery rate (FDR) was determined as described by Storey and Tibshirani [24].

**Candidate genes**

Based on a linkage disequilibrium (LD) decay below 0.2 within 3-6 kb in this *P. trichocarpa* population [20], a 6-kb upstream and downstream region was examined for candidate genes. The significant genomic regions must have at least one detectable SNP with a significance of -log10 p≥6 and be circumscribed by reproducible SNPs with a significance of -log10 p≥5 (Table S3). The underlying genes were identified based on the *P. trichocarpa* reference genome v3.0 in the Phytozome database using the BioMart tool ([https://phytozome.jgi.doe.gov](https://phytozome.jgi.doe.gov)) (Table S3). A complete gene list with InterPro descriptions was generated including loci of putative and/or unknown functions. Expression profiles of the gene models from various tissues: bud, leaf, various parts of root and stem, were obtained from the PhytoMine database ([https://phytozome.jgi.doe.gov/phytomine(begin).do](https://phytozome.jgi.doe.gov/phytomine/begin.do)).

**Results**

**Phenotyping for bark texture**

Phenotypic variation in the association population for bark texture ranged from smooth (score=1) to flaky bark (score=3) (Fig. S1). The mean values and standard errors for Clatskanie, Corvallis and Placerville sites were: CLA12 (mean: 1.53, SE: 0.70), CLA13 (mean: 1.50, SE: 0.69) and CLA15 (mean: 1.89, SE: 0.81), COR13 (mean: 1.71, SE: 0.62) and PLC14 (mean: 1.46, SE: 0.63), respectively. Phenotypic correlations were generally higher within the same environment with varying levels of significance across different environments (Table S1).
Genetic associations with phenotypic variation in bark texture

The GWAS using 8.2 million SNPs uncovered a total of 755 individual SNPs significantly (-log10 \( p \geq 5 \)) and reproducibly associated with bark texture distributed across all 19 chromosomes. The 175 genomic regions defined by these SNPs include 380 unique genes (Table S2). Specifically, 472, 255, 4, 28 and 17 SNPs were detected for the data sets Clatskanie 2012 (CLA12), CLA13, CLA15, Corvallis 2013 (COR13) and Placerville 2014 (PLC14), respectively. The Manhattan plots for the five location-by-year data sets with their replicates are shown in Figure 1.

At a significance threshold of -log10 \( p \geq 6 \) the 175 genomic regions were further reduced to 50 genomic regions ranging from 1-20 kb intervals, featuring 324 out of the 755 individual SNPs (Table S3). Nine genomic regions had no predicted genes within 50 kb of significant SNPs; accordingly, only 41 regions containing 98 out of the 380 genes were further used to describe gene annotations based on the interval of the significant genomic region with detectable SNPs (-log10 \( p \geq 5 \)).

Characterization of genomic regions

Table 1 summarizes the 41 genomic regions including SNPs with the highest significance value (detailed information in Table S3). These regions were categorized into two groups: reproducible across sites and reproducible across replicates within a site. Twenty-nine regions were reproducible across sites most of them having the exact same markers being significantly associated across multiple replicates/sites. The remaining twelve regions were reproducible across clonal replicates within a site. Comparing the significantly associated regions, they ranged from 1-20 kb with as few as 2 detected SNPs to as many as 53 SNPs. Most trait-associated SNP markers were located within noncoding regions (290) while 34 SNPs were within the coding regions of genes (nonsynonymous=26, synonymous=8). Details on all SNPs’ position and features are summarized in Table S3.

Seventeen regions had -log10 \( p \geq 7 \), with the highest significance on LG XV (8.44). Considering replicates, four regions, detected on LG VI, XV, XVI and XVII, were found in three different data sets (regions 1, 8, 11 and 12, Table 1). Four trait-associated
regions had more than 8 SNPs with -log10 p≥6 (region 5, 8, 11 and 31); whereas five regions contained 4-5 individual SNPs above -log10 p≥6 (region 6, 9, 17, 33 and 34, Table S3).

A detailed list of the underlying putative genes within the 41 regions and their interval is shown in Table S3. Some regions contained one gene while others included as many as nine, totaling 98 putative genes of which 75 were reproducible across sites and/or years and 23 were reproducible within sites/years. Analyzing the InterPro annotation and putative function of the significant genes along with the position/feature of SNPs, 32 of the 98 genes had unknown functions. Of the remaining genes several candidate genes had annotations possibly related to bark texture (Table 2). Eight of these genes occurred in QTL detected for bark texture in our previous study [16] (Fig. 2). Specifically, we found on LGI Potri.001G460900 (gene of unknown function), on LGVI the three genes Potri.006G173900 (gene of unknown function), Potri.006G200100 (gene of unknown function) and Potri.006G200200 (ring finger domain, ubiquitin ligase), on LGXII the three genes Potri.012G116800 (gene of unknown function), Potri.012G116900 and Potri.012G117000 (root hair defective 3), and finally on LGXVIII Potri.018G019900 (containing a NAF domain).

**Discussion**

The molecular machinery leading to alterations in bark texture is unknown. With the use of genome-wide association mapping we were able to link bark texture phenotypes to associated polymorphisms in the genome and identify putative genes related to bark development.

Even though the genes involved in bark texture variation are largely unknown, in the last two decades several studies have closely studied genes involved in related developmental processes. During radial growth mechanical stresses imposed in the phellogen create variation in bark roughness that can either produce a smooth or furrowed/flaky bark texture [25]. From studies on secondary growth from xylem mother cells (wood development), a comprehensive set of regulatory genes was discovered through various methods, including protein arrays linking genomics to proteomics and through EST sequencing providing expression analysis of genes during wood formation.
The study by Song et al. (2011) is one of the few which examined both xylem and phloem differentiation and created a subtracted gene list that included receptors, transporters, cell wall formation related and intracellular trafficking proteins [26].

Our results revealed multiple significant associations. Through GWAS, we were able to detect a number of putative genes involved in the control of bark texture, as well as differentiate between the genes responsible for the flaky bark texture in *P. trichocarpa* and the rough furrowed texture in *P. deltoides* in comparison with our previous work [16]. Overall the detected genomic regions involved in the control of bark texture in *P. trichocarpa* and *P. deltoides* vary. Eight genes out of 380 GWAS-identified genes were found to be co-locating with the QTL from the previous study, five of which have an unknown function. Two genes with annotated function, Potri.012G116900 and Potri.012G117000, on LG XII at Clatskanie (2012 and 2013) with a -log10 p=5.12, annotated as root hair defective 3, co-located with major highly reproductive QTL for bark texture [16]. The same QTL region was also associated with bark thickness and diameter, suggesting a role of this QTL in radial growth [16]. These genes putatively function as transmembrane proteins identified from the plasma membrane of *Populus* differentiating xylem and phloem [26].

Genome-wide association studies can suffer from too few observations, low repeatability and insufficient statistical thresholds. The large sample size in the present study (n=917) based on 8.2 million SNPs contributed to minimize the limitation of population size as well as defining the linkage disequilibrium rate between a SNP marker and a functional gene. Thus, most of significant associations were located within relatively small intervals (70% of genomic regions have ≤5 kb intervals). Among the 98 genes underlying the top 41 regions, a list of putative genes based on the highest p-values and reproducibility with possible functional control of bark texture were identified (Table 2).

*Radial growth and tissue differentiation*

During secondary growth, the vascular cambium consisting of meristematic tissues goes through cell division to radially generate secondary xylem to its inside and
secondary phloem to its outside [30]. The proportions of the cell types in inner bark (e.g., parenchyma, fibers, and sclereids) and the types of sieve elements, their sizes, and arrangements in the phloem vary among species, often forming patterns of taxonomic importance [31]. Vascular differentiation and differentiation of the outer bark involves cell division, orientated cell differentiation, cell expansion, cell wall thickening, and programmed cell death [26, 31]. Discontinuous periderms (textured bark) could result from variable radial meristematic activity in the cork cambium due to mechanical stresses from radial growth [25].

The gene Potri.015G124000, a membrane-associated apoptosis protein (NAP1), with a SNP located in its intron reached the highest significance score (-log10 p=8.44) and reproducibility across three datasets, CLA12, CLA13 and PLC14 (Table 2). It is an orthologous to the *A. thaliana* gene AT2G35110, NCK-associated protein (NAP1), an adaptor protein thought to modulate actin nucleation. NAP1 is a component of the WAVE regulatory protein complex that acts as an activator of the ARP2/3 complex involved in actin nucleation and trichome morphogenesis. Mutants in this gene display distorted trichomes and irregularities in trichome branch positioning and expansion [32-34]. NAP1 is thus involved in plant cell morphogenesis and plays an essential role in cell expansion.

Another gene family, Cytochromes, have been detected within six genomic regions (four having (-log10 p>7) across all years for Clatskanie and Corvallis sites (Table 2). Cytochromes are brassinosteroids perceived by the plasma membrane-localized leucine-rich repeat-receptor kinase BRI1 with essential roles in developmental processes including cell expansion and vascular differentiation [35]. For instance, loss-of-function of *brl1* causes abnormal phloem:xylem differentiation ratios and vascular defects [36]. Other studies have revealed that Cytochrome P450 plays a functional role in secondary metabolite biosynthesis, specifically in lignin monomer composition and controlling cell elongation and de-etiolation in *Arabidopsis* [37, 38]. While an overexpression of F5H, a cytochrome P450-dependent monooxygenases, in *Arabidopsis* led to the accumulation of lignin indicating a key regulatory role in syringyl lignin biosynthesis [37].
Several additional GWAS candidate genes, including pectate lyase, have been reported to regulate cell wall loosening and expansion [39]. Multicopper oxidase (Potri.019G088500, ortholog AT4G28250), known as an Expansin protein, functioning in abscission and cell wall loosening, has a significant SNP within the coding sequence (-log10 p=7.26) resulting in a nonsynonymous substitution and was detected in the CLA12 dataset [40].

**Suberin accumulation**

Suberization occurs during the development of the secondary radial meristems, i.e., the cork cambium [3]. Phellem is multilayered dead tissue that is made impervious by the disposition of suberin onto their cell walls, a waxy substance that is highly impermeable to gases and water to provide the internal cells of the plants with extra insulation and protection [2, 3].

The genes Potri.017G034700 and Potri.017G035400, wall-associated receptor kinases (PR5K-like), are found within highly significant genomic regions (-log10 p=7.90, synonymous SNP in coding sequence; -log10 p=5.72, SNP in intron, respectively) and are reproducible across year 2012 and 2013 in Clatskanie (Table 2). They are related to a family of protein-serine/threonine kinases and involved in self-incompatibility and disease resistance and are orthologous to the *Arabidopsis* gene AT3G55550 [41, 42]. A highly significant synonymous SNP (-log10 p=7.13) was found in the coding region of the gene Potri.017G034500, an ortholog to the same *Arabidopsis* gene AT3G55550, however it was only characterized as *Populus* gene with protein kinase domain of unknown family (Table 2). Interestingly, an ortholog of the PR5K-like gene, AT5G38280, was identified as a candidate gene involved in apple russetting [43], which results from the accumulation of suberin on the inner part of the cell wall of the outer epidermal cell layers.

Other regulatory genes for suberin accumulation are the MYB transcription factors of which three Myb-like DNA-binding protein genes were detected in Clatskanie across all years at -log10 p=7.49 (Table 2). The MYB gene family have been reported as key regulators in the biosynthesis of lignin and the regulation of secondary cell wall formation [44-47]. In *Arabidopsis*, MYB46 and MYB83 are direct targets of a group of
NAC domain master regulators controlling secondary wall biosynthesis [47-49]. In addition, in *Populus*, several genes were found to be involved in the regulation of wood formation, including PrtMYB3 and PrtMYB20, while a NAC protein Potri.006G277000, detected in the COR13 dataset, is involved in the biosynthesis and regulation of *Populus* suberin affecting the bark periderm, specifically the cork (phellem) [43, 50-52].

*Programmed cell death*

Programmed cell death is important in secondary plant development and is involved in the generation of the vascular system [53]. Programmed cell death also contributes to the development of bark and its abscission [54]. Abscission results in the separation of the bark tissue from the trunk without injury to living tissue. Bark abscission creates a protection layer of the newly exposed tissue [54]. Lignification occurs in xylem cells during wood development and in bark [4, 55, 56]. When lignification is completed, vessel elements undergo programmed cell death, which involves the hydrolysis of the protoplast [4, 57, 58]. Nonetheless, the specific effects of cell death on bark has not been studied.

Six genes with ring finger domain and ubiquitin ligase function were detected across all sites and several years, of which Potri.006G200200 co-locates with a QTL cluster on LG VI associated with bark texture from our previous study [16]. The genes Potri.001G450700 and Potri.014G138000 (orthologs AT3G61460 and AT1G63900, respectively) have the highest significant values at -log10 p=7.75 and 7.56, respectively, and their putative functions are characterized as SFC and E3 ubiquitin ligases, respectively (Table 2). E3 ubiquitin ligases are known as positive regulators of cell death and defense across various plants [59]. Studies have suggested their involvement in regulation of cellular elongation in *Arabidopsis* and in the biosynthesis of tension wood in *Populus* [29, 60]. The ubiquitin-dependent pathway of proteolysis is essential for vascular development and overexpression of a mutant form of ubiquitin unable to form polyubiquitin chains in transgenic tobacco resulted in abnormal vascular development [61].

One way for a plant to defend itself is by using molecular switches of plant disease resistance, such as NB-ARC domain containing proteins, and activating a cascade
leading to cell death [62, 63]. Four genes within several genomic regions (-log10 p≤7.13) detected across all three locations and years contain a NB-ARC (nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4) domain, of which two had SNPs within genes, one in the coding sequence (synonymous substitution) and the other in the genes’ intron (Table 2). A recent study by Fossdal et al. examined gene expression of secondary phloem of Norway spruce after it had been wounded and inoculated with the fungus Ceratocystis polonica [64]. NB resistance genes showed a significant increase in expression along with five miRNAs with putative NB targets [64].

Although seven Populus genes were characterized as kinases with unknown family, two genes, Potri.015G124100 and Potri.017G034100 (-log10 p=8.44 and 7.13, respectively), were identified across three of our test sites and are orthologous to Arabidopsis genes are known as cysteine-rich receptor-like protein kinases (CRKs). Induced expression of these CRKs triggered hypersensitive response-like cell death in transgenic plants [65].

**Future Work**

While the candidate gene list in Table 2 provides many promising putative genes involved in cell wall differentiation they need to be further studied to validate their involvement in bark biosynthesis through transgenic lines and forward genetic methods. Genes of unknown function detected within highly significant regions open new opportunities in the future and should be functionally characterized. Two genes, Potri.006G008100 and Potri.017G035600 (-log10 p=6.91 and 7.73, respectively), are top candidates for functional characterization along with 12 other genes with high significance (-log10 p≥7) (Table S3).
**Conclusions**

The combination of QTL and GWAS mapping supports the identification of candidate genes. QTL mapping increases the ability to reduce false positive rates whereas GWAS has a much higher resolution than QTL mapping and can narrow down the genomic region to individual candidate genes [66-68]. In our case, the combination of QTL and GWAS mapping provided insights on the different genes responsible for the variations in bark texture ranging from smooth and furrowed across *Populus* species.
References


**Figures and Tables**

**Figure 1** Manhattan plots for GWAS for the *P. trichocarpa* BESC population. Standard −log10 P plots of the study results. The five datasets across three sites: Clatskanie (CLA) for year 2012, 2013 and 2015 (a), Corvallis (COR) for year 2013 (b) and Placerville (PLC) for year 2014 (b); BT1, BT2, BT3: bark texture replicate 1, 2, 3.
Figure 2 Location of top candidate genes on the linkage map derived from a cross of a female hybrid clone (*Populus trichocarpa x P. deltoides*) with *P. deltoides* (Bdeir et al. 2017). The linkage map is anchored to the genome of *Populus trichocarpa* (V3 assembly, Kelleher et al. 2007), see Table 1. QTL intervals for bark texture are shown in yellow based on Bdeir et al. (2017). Vertical lines show the location of top candidate genes on the *P. trichocarpa* genome detected near significant and highly reproducible SNPs in the genome-wide association mapping analysis. Red, green and blue genes are putatively involved in radial growth and tissue differentiation, suberin accumulation and programmed cell death, respectively.
Table 1 The 41 genomic regions with reproducible SNPs most significantly associated with bark texture in the genome-wide association study (GWAS).

<table>
<thead>
<tr>
<th>Region #</th>
<th>LG</th>
<th>SNP position interval (kb)</th>
<th>Interval (kb)</th>
<th>Population</th>
<th>SNP position</th>
<th>Corr -log10 P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>XV</td>
<td>13756415-13759891</td>
<td>3</td>
<td>PLCV14_BT2</td>
<td>13,759,891</td>
<td>8.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12_BT2</td>
<td>13,756,415</td>
<td>5.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT2</td>
<td>13,756,415</td>
<td>5.22</td>
</tr>
<tr>
<td>2</td>
<td>XVII</td>
<td>2930303-2942559</td>
<td>12</td>
<td>CLA12_BT3</td>
<td>2,931,167</td>
<td>7.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12cookieBT1</td>
<td>2,931,167</td>
<td>6.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT1</td>
<td>2,941,148</td>
<td>5.07</td>
</tr>
<tr>
<td>3</td>
<td>V</td>
<td>16,609,078</td>
<td>1</td>
<td>CLA13_BT1</td>
<td>16,609,078</td>
<td>7.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12_BT1</td>
<td>16,609,078</td>
<td>7.74</td>
</tr>
<tr>
<td>4</td>
<td>XIV</td>
<td>10,543,886</td>
<td>1</td>
<td>CLA12_BT2</td>
<td>10,543,886</td>
<td>7.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT2</td>
<td>10,543,886</td>
<td>6.12</td>
</tr>
<tr>
<td>5</td>
<td>I</td>
<td>48502337-48513466</td>
<td>11</td>
<td>CLA12_BT1</td>
<td>48,504,159</td>
<td>7.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12_BT3</td>
<td>48,504,159</td>
<td>5.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT2</td>
<td>48,504,159</td>
<td>5.68</td>
</tr>
<tr>
<td>6</td>
<td>XIV</td>
<td>3,085,288</td>
<td>1</td>
<td>CLA12_BT3</td>
<td>3,085,288</td>
<td>7.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12_BT1</td>
<td>3,085,288</td>
<td>7.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12cookieBT1</td>
<td>3,085,288</td>
<td>6.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT1</td>
<td>3,085,288</td>
<td>6.48</td>
</tr>
<tr>
<td>7</td>
<td>I</td>
<td>44916073-44932416</td>
<td>16</td>
<td>CLA12_BT2</td>
<td>44,931,643</td>
<td>7.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT2</td>
<td>44,931,643</td>
<td>6.44</td>
</tr>
<tr>
<td>8</td>
<td>XVII</td>
<td>2894584-2915048</td>
<td>20</td>
<td>CLA12_BT3</td>
<td>2,905,659</td>
<td>7.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12cookieBT1</td>
<td>2,905,659</td>
<td>6.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>VIII</td>
<td>11209007-11217624</td>
<td>9</td>
<td>CLA13_BT1</td>
<td>11,209,437</td>
<td>7.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT1</td>
<td>11,209,437</td>
<td>5.95</td>
</tr>
<tr>
<td>10</td>
<td>XI</td>
<td>11,521,810</td>
<td>1</td>
<td>CLA13_BT1</td>
<td>11,521,810</td>
<td>7.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT1</td>
<td>11,521,810</td>
<td>6.55</td>
</tr>
<tr>
<td>11</td>
<td>VI</td>
<td>21529197-21533794</td>
<td>5</td>
<td>CLA12_BT2</td>
<td>21,530,675</td>
<td>7.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT2</td>
<td>21,530,675</td>
<td>6.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA15</td>
<td>21,530,272</td>
<td>5.24</td>
</tr>
<tr>
<td>12</td>
<td>XVI</td>
<td>5614771-5630821</td>
<td>16</td>
<td>CLA12_BT2</td>
<td>5,626,773</td>
<td>6.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT2</td>
<td>5,626,773</td>
<td>5.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PLCV14_BT1</td>
<td>5,628,967</td>
<td>5.77</td>
</tr>
<tr>
<td>13</td>
<td>VI</td>
<td>543,317</td>
<td>1</td>
<td>CLA12_BT2</td>
<td>543,317</td>
<td>6.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT2</td>
<td>543,317</td>
<td>6.80</td>
</tr>
<tr>
<td>14</td>
<td>X</td>
<td>8131399-8141525</td>
<td>10</td>
<td>CLA12cookieBT1</td>
<td>8,131,399</td>
<td>6.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12_BT3</td>
<td>8,131,399</td>
<td>6.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT1</td>
<td>8,141,525</td>
<td>5.33</td>
</tr>
<tr>
<td>15</td>
<td>VII</td>
<td>16608390-16609078</td>
<td>1</td>
<td>CLA13_BT2</td>
<td>11,609,206</td>
<td>6.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12_BT2</td>
<td>11,609,206</td>
<td>5.40</td>
</tr>
<tr>
<td>16</td>
<td>IV</td>
<td>21617052-21619943</td>
<td>3</td>
<td>CLA13_BT2</td>
<td>21,619,943</td>
<td>6.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12_BT1</td>
<td>21,619,943</td>
<td>5.86</td>
</tr>
<tr>
<td>17</td>
<td>V</td>
<td>21642667-21646857</td>
<td>4</td>
<td>CLA12_BT3</td>
<td>21,644,588</td>
<td>6.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>XVII</td>
<td>15160420-15163613</td>
<td>3</td>
<td>CLA13_BT1</td>
<td>15,160,420</td>
<td>6.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA2_BT1</td>
<td>15,160,420</td>
<td>5.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>XVIII</td>
<td>6040783-6046262</td>
<td>6</td>
<td>CLA12_BT1</td>
<td>6,042,491</td>
<td>6.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT1</td>
<td>6,042,491</td>
<td>5.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>IV</td>
<td>8874775-8883033</td>
<td>8</td>
<td>CLA12_BT2</td>
<td>8,877,397</td>
<td>6.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT2</td>
<td>8,877,397</td>
<td>5.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>XV</td>
<td>3,872,558</td>
<td>1</td>
<td>CLA12_BT1</td>
<td>3,872,558</td>
<td>6.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT1</td>
<td>3,872,558</td>
<td>5.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>XIV</td>
<td>7,277,970</td>
<td>1</td>
<td>CLA12_BT2</td>
<td>7,277,970</td>
<td>6.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT2</td>
<td>7,277,970</td>
<td>5.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>V</td>
<td>2,047,117</td>
<td>1</td>
<td>CLA13_BT1</td>
<td>2,047,117</td>
<td>6.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12_BT1</td>
<td>2,047,117</td>
<td>5.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>XVI</td>
<td>889,154</td>
<td>1</td>
<td>CLA12_BT1</td>
<td>889,154</td>
<td>6.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA13_BT1</td>
<td>889,154</td>
<td>5.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>XII</td>
<td>2,115,668</td>
<td>1</td>
<td>CLA13_BT2</td>
<td>2,115,668</td>
<td>6.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12_BT2</td>
<td>2,115,668</td>
<td>6.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>XI</td>
<td>15754964-15758609</td>
<td>4</td>
<td>CLA13_BT1</td>
<td>15,758,609</td>
<td>6.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12_BT2</td>
<td>15,754,964</td>
<td>5.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>XIII</td>
<td>1,649,986</td>
<td>1</td>
<td>CLA13_BT1</td>
<td>1,649,986</td>
<td>6.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CLA12_BT1</td>
<td>1,649,986</td>
<td>5.57</td>
</tr>
<tr>
<td>Line</td>
<td>Page</td>
<td>Reference</td>
<td>Count</td>
<td>Series</td>
<td>Sequence</td>
<td>Value</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>-----------</td>
<td>-------</td>
<td>--------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td>29</td>
<td>VIII</td>
<td>10028445-10032564</td>
<td>4</td>
<td>CLA13_BT2</td>
<td>CLA12cookieBT1</td>
<td>10,028,892</td>
</tr>
<tr>
<td>30</td>
<td>I</td>
<td>31275502-31277465</td>
<td>2</td>
<td>CLA12_BT2</td>
<td>CLA12cookieBT1</td>
<td>31,276,103</td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>3001881-3008293</td>
<td>7</td>
<td>CLA12cookieBT1</td>
<td>CLA12_BT3</td>
<td>3,007,903</td>
</tr>
<tr>
<td>32</td>
<td>VII</td>
<td>10,801,084</td>
<td>1</td>
<td>CLA12cookieBT1</td>
<td>CLA12_BT3</td>
<td>10,801,084</td>
</tr>
<tr>
<td>33</td>
<td>XIX</td>
<td>4752983-4762402</td>
<td>10</td>
<td>CLA12_BT3</td>
<td>CLA12cookieBT1</td>
<td>4,762,402</td>
</tr>
<tr>
<td>34</td>
<td>XI</td>
<td>10996808-10996815</td>
<td>1</td>
<td>CLA12_BT3</td>
<td>CLA12cookieBT1</td>
<td>10,996,815</td>
</tr>
<tr>
<td>35</td>
<td>XIX</td>
<td>12020500-12025588</td>
<td>5</td>
<td>CLA12_BT3</td>
<td>CLA12cookieBT1</td>
<td>12,025,588</td>
</tr>
<tr>
<td>36</td>
<td>VI</td>
<td>13,299,251</td>
<td>1</td>
<td>CLA12cookieBT1</td>
<td>CLA12_BT3</td>
<td>13,299,251</td>
</tr>
<tr>
<td>37</td>
<td>XVI</td>
<td>12838532-12843036</td>
<td>5</td>
<td>COR13_BT2</td>
<td>COR13_BT1</td>
<td>12,838,536</td>
</tr>
<tr>
<td>38</td>
<td>XVI</td>
<td>2627146-2630738</td>
<td>3</td>
<td>COR13_BT3</td>
<td>COR13_BT2</td>
<td>2,627,146</td>
</tr>
<tr>
<td>39</td>
<td>XI</td>
<td>1,779,913</td>
<td>1</td>
<td>CLA12cookieBT1</td>
<td>CLA12_BT3</td>
<td>1,779,913</td>
</tr>
<tr>
<td>40</td>
<td>II</td>
<td>15467058-15489644</td>
<td>22</td>
<td>COR13_BT2</td>
<td></td>
<td>15,469,012</td>
</tr>
</tbody>
</table>
Populations: the name indicates site, year and replicate (e.g. CLA12_BT3, Clatskanie, year 2012, bark texture replicate three). The first 29 genomic regions with highly significant SNPs are found reproducible across sites, while the remaining 12 regions are reproducible within a site. For each genomic region and experiment the most significant associations are shown. Highly significant SNPs are in bold (-log10 P ≥ 6). The interval in which significant SNPs were detected at a significance threshold of -log10 P ≥ 5, the populations and replicates in which SNPs were detected in corresponding genomic regions along with the highest corrected p-value are shown. Detailed information on SNPs and their genomic location can be found in Table S4.

<table>
<thead>
<tr>
<th>LG</th>
<th>Population</th>
<th>Start</th>
<th>End</th>
<th>Corrected p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>XVII</td>
<td>41 COR13_BT1</td>
<td>15,482,878</td>
<td>5.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CLA12cookieBT1</td>
<td>9,570,806</td>
<td>6.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CLA12_BT3</td>
<td>9,570,806</td>
<td>5.97</td>
<td></td>
</tr>
</tbody>
</table>
Table 2 Genomic location and putative functions SNP containing genes and of genes adjacent to SNPs associated with bark texture identified by genome-wide association study (GWAS).

<table>
<thead>
<tr>
<th>Region #</th>
<th>Population</th>
<th>Highest - log10 P</th>
<th>SNP within gene feature</th>
<th>Gene Name</th>
<th>Putative function</th>
<th>Arabidopsis ortholog</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CLA12, CLA13, PLC14</td>
<td>8.44</td>
<td>Intron</td>
<td>Potri.015G124000</td>
<td>Membrane-associated apoptosis protein, NAP1</td>
<td>AT2G35110</td>
</tr>
<tr>
<td>1</td>
<td>CLA12, CLA13, PLC14</td>
<td>8.44</td>
<td>Intron, Upstream/Downstream</td>
<td>Potri.015G124100</td>
<td>Protein kinase domain</td>
<td>AT4G11480</td>
</tr>
<tr>
<td>31</td>
<td>CLA12</td>
<td>7.74</td>
<td>CDS, NS, Downstream</td>
<td>Potri.017G035500</td>
<td>Protein kinase domain</td>
<td>AT5G38260</td>
</tr>
<tr>
<td>8</td>
<td>CLA12, CLA13, PLC14</td>
<td>7.13</td>
<td>CDS, S, Upstream/downstream</td>
<td>Potri.017G034500</td>
<td>Protein kinase domain</td>
<td>AT3G55550</td>
</tr>
<tr>
<td>8</td>
<td>CLA12, CLA13, PLC14</td>
<td>7.13</td>
<td>Upstream/Downstream</td>
<td>Potri.017G034100</td>
<td>Protein tyrosine kinase</td>
<td>AT4G04490</td>
</tr>
<tr>
<td>20</td>
<td>PLC14, COR13</td>
<td>6.34</td>
<td>Downstream</td>
<td>Potri.006G279900</td>
<td>Protein tyrosine kinase</td>
<td>AT4G31170</td>
</tr>
<tr>
<td>20</td>
<td>PLC14, COR13</td>
<td>6.34</td>
<td>3' UTR</td>
<td>Potri.006G280000</td>
<td>Protein tyrosine kinase</td>
<td>AT2G24360</td>
</tr>
<tr>
<td>NA</td>
<td>COR13, CLA15</td>
<td>5.87</td>
<td>Downstream, intergenic</td>
<td>Potri.014G161400</td>
<td>Protein kinase domain</td>
<td>AT5G21222</td>
</tr>
<tr>
<td>2</td>
<td>CLA12, CLA13</td>
<td>7.90</td>
<td>CDS, S, Upstream/Downstream</td>
<td>Potri.017G034700</td>
<td>Wall-associated receptor kinase galacturonan-binding</td>
<td>None</td>
</tr>
<tr>
<td>NA</td>
<td>CLA12, CLA13</td>
<td>5.73</td>
<td>Intron</td>
<td>Potri.017G035400</td>
<td>Wall-associated receptor kinase galacturonan-binding</td>
<td>AT3G55550</td>
</tr>
<tr>
<td>3</td>
<td>CLA12, CLA13</td>
<td>7.86</td>
<td>Intron</td>
<td>Potri.005G162300</td>
<td>VQ motif</td>
<td>AT1G78310</td>
</tr>
<tr>
<td>4</td>
<td>CLA12, CLA13</td>
<td>7.75</td>
<td>Upstream, intergenic</td>
<td>Potri.014G138000</td>
<td>Zinc finger, C3HC4 type (RING finger), ubiquitin ligase</td>
<td>AT1G63900</td>
</tr>
<tr>
<td></td>
<td>Gene IDs</td>
<td>Log2 Fold Change</td>
<td>Location Type</td>
<td>Gene Accession</td>
<td>Domain Description</td>
<td>Target Genes</td>
</tr>
<tr>
<td>---</td>
<td>----------</td>
<td>-----------------</td>
<td>---------------</td>
<td>----------------</td>
<td>--------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>5</td>
<td>CLA12, CLA13</td>
<td>7.56</td>
<td>Upstream, intergenic</td>
<td>Potri.001G450700</td>
<td>Ring finger domain, ubiquitin ligase</td>
<td>None</td>
</tr>
<tr>
<td>10</td>
<td>CLA12, CLA13</td>
<td>7.03</td>
<td>Upstream, intergenic</td>
<td>Potri.011G094800</td>
<td>Ring finger domain</td>
<td>AT5G55970</td>
</tr>
<tr>
<td>11</td>
<td>CLA12, CLA13, CLA15</td>
<td>7.00</td>
<td>Downstream, intergenic</td>
<td>Potri.006G200200</td>
<td>Ring finger domain, ubiquitin ligase</td>
<td>AT5G22000</td>
</tr>
<tr>
<td>NA</td>
<td>PLC14, COR13</td>
<td>5.78</td>
<td>Intron</td>
<td>Potri.019G060900</td>
<td>Zinc finger, C3HC4 type (RING finger), DNA helicase</td>
<td>AT2G40770</td>
</tr>
<tr>
<td>NA</td>
<td>CLA12, PLC14</td>
<td>5.26</td>
<td>Upstream, intergenic</td>
<td>Potri.009G080500</td>
<td>RING-type zinc-finger, ubiquitin ligase</td>
<td>None</td>
</tr>
<tr>
<td>5</td>
<td>CLA12, CLA13</td>
<td>7.56</td>
<td>Intron</td>
<td>Potri.001G450800</td>
<td><strong>Ribosomal protein L11, RNA binding domain</strong></td>
<td>AT1G32990</td>
</tr>
<tr>
<td>6</td>
<td>CLA12, CLA13</td>
<td>7.51</td>
<td>Upstream, intergenic</td>
<td>Potri.014G037800</td>
<td>Cytochrome P450</td>
<td>AT5G25180</td>
</tr>
<tr>
<td>32</td>
<td>CLA12</td>
<td>7.47</td>
<td>Upstream, intergenic</td>
<td>Potri.007G083300</td>
<td>Cytochrome P450</td>
<td>AT2G24180</td>
</tr>
<tr>
<td>7</td>
<td>CLA12, CLA13</td>
<td>7.41</td>
<td>Upstream, intergenic</td>
<td>Potri.001G424300</td>
<td>Cytochrome P450</td>
<td>None</td>
</tr>
<tr>
<td>7</td>
<td>CLA12, CLA13</td>
<td>7.41</td>
<td>Downstream, intergenic</td>
<td>Potri.001G424400</td>
<td>Cytochrome P450</td>
<td>None</td>
</tr>
<tr>
<td>NA</td>
<td>COR13, CLA15</td>
<td>5.87</td>
<td>Upstream, intergenic</td>
<td>Potri.014G161700</td>
<td>Eukaryotic cytochrome b561</td>
<td>None</td>
</tr>
<tr>
<td>NA</td>
<td>CLA12</td>
<td>5.35</td>
<td>Downstream, intergenic</td>
<td>Potri.007G083700</td>
<td>Cytochrome P450</td>
<td>AT2G24180</td>
</tr>
<tr>
<td>30</td>
<td>CLA12</td>
<td>7.49</td>
<td>Downstream, intergenic</td>
<td>Potri.001G309100</td>
<td>Myb/SANT-like DNA-binding domain</td>
<td>AT1G76880</td>
</tr>
<tr>
<td>NA</td>
<td>CLA12, CLA13</td>
<td>5.45</td>
<td>Downstream, intergenic</td>
<td>Potri.003G200000</td>
<td>Myb/SANT-like DNA-binding domain</td>
<td>AT3G14180</td>
</tr>
<tr>
<td>NA</td>
<td>CLA12</td>
<td>5.98</td>
<td>Downstream, intergenic</td>
<td>Potri.006G085900</td>
<td>Myb-like DNA-binding domain</td>
<td>AT2G37630</td>
</tr>
<tr>
<td>Rank</td>
<td>Reference</td>
<td>Gene</td>
<td>Description</td>
<td>Domain</td>
<td>AT Accession</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>------</td>
<td>-------------</td>
<td>--------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>7.41</td>
<td>CLA12, CLA13</td>
<td>Intron</td>
<td>Potri.001G424600</td>
<td>EAP30/Vps36 family, RNA polymerase II transcription factor complex</td>
<td>AT4G27040</td>
<td></td>
</tr>
<tr>
<td>7.41</td>
<td>CLA12, CLA13</td>
<td>5' UTR</td>
<td>Potri.001G424700</td>
<td>Kua-ubiquitin conjugating enzyme hybrid localisation domain,</td>
<td>AT4G27030</td>
<td></td>
</tr>
<tr>
<td>7.26</td>
<td>CLA12</td>
<td>CDS, NS</td>
<td>Potri.019G088500</td>
<td>Multicopper oxidase</td>
<td>AT5G58910</td>
<td></td>
</tr>
<tr>
<td>7.13</td>
<td>CLA12, CLA13, PLC14</td>
<td>Upstream/Downstream</td>
<td>Potri.017G034400</td>
<td>NB-ARC domain, Leucine rich repeat</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>6.50</td>
<td>CLA12, CLA13, COR13</td>
<td>CDS, NS</td>
<td>Potri.017G144100</td>
<td>NB-ARC domain, Leucine rich repeat</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CLA12, CLA13</td>
<td>Downstream</td>
<td>Potri.006G273900</td>
<td>NB-ARC domain, Leucine rich repeat</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CLA12, CLA13</td>
<td>Intron</td>
<td>Potri.006G274000</td>
<td>NB-ARC domain, Leucine rich repeat</td>
<td>AT5G06040</td>
<td></td>
</tr>
<tr>
<td>7.13</td>
<td>CLA12, CLA13, PLC14</td>
<td>CDS, S and NS, Upstream/Downstream</td>
<td>Potri.017G034200</td>
<td>Leucine rich repeat</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>6.47</td>
<td>CLA12, CLA13</td>
<td>Upstream, intergenic</td>
<td>Potri.018G057100</td>
<td>Leucine rich repeat N-terminal domain</td>
<td>AT5G25930</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CLA12, CLA13</td>
<td>Downstream, intergenic</td>
<td>Potri.017G035300</td>
<td>Leucine rich repeat, Apoptotic ATPase</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CLA12, CLA13</td>
<td>Downstream, intergenic</td>
<td>Potri.003G027200</td>
<td>Leucine Rich Repeat</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>CLA12, CLA13</td>
<td>Upstream, intergenic</td>
<td>Potri.002G019900</td>
<td>Leucine rich repeat N-terminal domain</td>
<td>AT1G75820</td>
<td></td>
</tr>
<tr>
<td>7.13</td>
<td>CLA12, CLA13, PLC14</td>
<td>CDS, S and NS, Downstream</td>
<td>Potri.017G033900</td>
<td>BED zinc finger, Leucine rich repeat, Apoptotic ATPase</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>7s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>8</td>
<td>CLA12, CLA13, PLC14</td>
<td>7.13</td>
<td>Downstream</td>
<td>Potri.017G033800</td>
<td>Potential DNA-binding domain</td>
<td>None</td>
</tr>
<tr>
<td>10</td>
<td>CLA12, CLA13</td>
<td>7.03</td>
<td>Downstream, intergenic</td>
<td>Potri.011G094700</td>
<td>ABC transporter</td>
<td>AT5G60740</td>
</tr>
<tr>
<td>NA</td>
<td>CLA12</td>
<td>5.98</td>
<td>Upstream, intergenic</td>
<td>Potri.006G086000</td>
<td>ABC transporter</td>
<td>AT5G02270</td>
</tr>
<tr>
<td>NA</td>
<td>COR13</td>
<td>5.69</td>
<td>Downstream, intergenic</td>
<td>Potri.006G276600</td>
<td>ABC1 family</td>
<td>AT5G24810</td>
</tr>
<tr>
<td>12</td>
<td>CLA12, CLA13, PLC14</td>
<td>6.99</td>
<td>Downstream, intergenic</td>
<td>Potri.016G074200</td>
<td>TCP family transcription factor</td>
<td>None</td>
</tr>
<tr>
<td>12</td>
<td>CLA12, CLA13, PLC14</td>
<td>6.99</td>
<td>Downstream, intergenic</td>
<td>Potri.016G074400</td>
<td>PROTEASE S28 PRO-X CARBOXYPEPTIDASE-RELATED</td>
<td>None</td>
</tr>
<tr>
<td>13</td>
<td>CLA12, CLA13</td>
<td>6.92</td>
<td>Upstream</td>
<td>Potri.006G008200</td>
<td>X8 domain</td>
<td>None</td>
</tr>
<tr>
<td>16</td>
<td>CLA12, CLA13</td>
<td>6.73</td>
<td>Intron and 3' UTR</td>
<td>Potri.004G207300</td>
<td>EF-hand domain pair</td>
<td>None</td>
</tr>
<tr>
<td>14</td>
<td>CLA12, CLA13</td>
<td>6.75</td>
<td>Upstream/Downstream, intergenic</td>
<td>Potri.010G050900</td>
<td>WD domain, G-beta repeat</td>
<td>AT1G24530</td>
</tr>
<tr>
<td>36</td>
<td>CLA12</td>
<td>6.64</td>
<td>Intron</td>
<td>Potri.006G151800</td>
<td>WD domain, G-beta repeat</td>
<td>AT2G19520</td>
</tr>
<tr>
<td>22</td>
<td>CLA12, CLA13</td>
<td>6.28</td>
<td>Upstream, intergenic</td>
<td>Potri.015G041200</td>
<td>WD domain, G-beta repeat</td>
<td>AT3G18060</td>
</tr>
<tr>
<td>16</td>
<td>CLA12, CLA13</td>
<td>6.73</td>
<td>5' UTR and Intron</td>
<td>Potri.004G207400</td>
<td>PEPTIDASE M20 FAMILY MEMBER</td>
<td>AT4G38225</td>
</tr>
<tr>
<td>20</td>
<td>PLC14, COR13</td>
<td>6.34</td>
<td>CDA, NS and Intron</td>
<td>Potri.006G280200</td>
<td>VPRBP PROTEIN-RELATED</td>
<td>AT4G31160</td>
</tr>
<tr>
<td>20</td>
<td>PLC14, COR13</td>
<td>6.34</td>
<td>CDS, NS</td>
<td>Potri.006G280100</td>
<td>Amino acid kinase family</td>
<td>AT3G57560</td>
</tr>
<tr>
<td></td>
<td>Genes</td>
<td></td>
<td>Description</td>
<td></td>
<td>Genomic Location</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------</td>
<td>---</td>
<td>------------------------------------------------</td>
<td>---</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>CLA12, CLA13</td>
<td>6.27</td>
<td>Intron</td>
<td>Potri.014G093300</td>
<td><strong>Phosphatidylinositol-4-phosphate 5-Kinase</strong> AT2G26420</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>CLA12</td>
<td>6.14</td>
<td>Intron</td>
<td>Potri.011G020700</td>
<td><strong>Ankyrin repeats (3 copies), 26S proteasome regulatory complex, subunit PSMD10</strong> None</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>CLA12, CLA13</td>
<td>6.13</td>
<td>Downstream, intergenic</td>
<td>Potri.011G132700</td>
<td><strong>Breast carcinoma amplified sequence 3</strong> AT3G20720</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>CLA12, CLA13</td>
<td>6.13</td>
<td>Upstream, intergenic</td>
<td>Potri.011G132800</td>
<td><strong>N-terminal region of Chorein, a TM vesicle-mediated sorter</strong> AT5G54730</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>CLA12, CLA13</td>
<td>6.01</td>
<td>Upstream, intergenic</td>
<td>Potri.008G148800</td>
<td><strong>Pectate lyase</strong> AT3G01270</td>
<td></td>
</tr>
</tbody>
</table>

Genes are sorted according to highest -log10 values within family/description. Sites: CLA: Clatskanie, COR: Corvallis, PLC: Placerville. The putative function of SNP containing genes is shown in bold. Populations: the name indicates site and year. CDS: coding sequences, S: synonymous, NS: nonsynonymous. NA: not applicable since these reproducible regions (-log10 of ≥ 5) are not part of the 41 highly significant regions (-log10 of ≥ 6) but these genes from the same putative family/description are added. The genomic location of genes is given by the gene name (e.g., Potri.015G124000 is located on linkage group 15). Detailed information can be found in Table S4.
Supplementary Figure 1 Bark texture scale from smooth (1), medium flaky (2) to rough flaky (3) for the *P. trichocarpa* populations.

Supplementary Figure 2 Frequency distribution for bark texture across Clatskanie (CLA), Corvallis (COR) and Placerville (PLC) sites and various years (a, b, c, respectively).
### Supplementary Table 1  
Spearman’s rank correlation coefficients for bark texture phenotypes across five datasets.

<table>
<thead>
<tr>
<th></th>
<th>Replicate</th>
<th>COR13</th>
<th>CLA12</th>
<th>CLA13</th>
<th>CLA15</th>
<th>PLC14</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>COR13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLA12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>0.3523</td>
<td></td>
<td>x</td>
<td>0.4116</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.4048</td>
<td>0.3740</td>
<td>0.4893</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.4080</td>
<td>0.4172</td>
<td>0.4383</td>
<td>0.9051</td>
<td>x</td>
</tr>
<tr>
<td>CLA13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>0.3842</td>
<td>0.3541</td>
<td>0.4956</td>
<td>0.9581</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.3842</td>
<td>0.3672</td>
<td>0.4025</td>
<td>0.7822</td>
<td>0.9783</td>
</tr>
<tr>
<td>CLA15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>-0.3775</td>
<td>-0.4129</td>
<td>-0.5250</td>
<td>-0.4485*</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>0.0431ns</td>
<td>0.2897</td>
<td>0.1122ns</td>
<td>0.3098*</td>
</tr>
<tr>
<td>PLC14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Three different sites including Corvallis (COR), Clatskanie (CLA) and Placerville (PLC), of which CLA datasets were collected across three years. Statistically significant correlations across different environments are shown in bold ($p<0.0001$), * $p<0.05$, ns = not significant.
Supplementary Table 2 All 1825 SNPs associated with bark texture and the underlying genes within a 6 kb interval. The population in which these SNP were detected in, their position, linkage group and significance score along with the physical localization, annotation and expression profiles of the underlying genes are summarized.

Supplementary Table 3 The 50 genomic regions most strongly associated with bark texture and the underlying genes. The population in which these 129 individual SNPs were detected, position, linkage group and significance score as well as the physical localization and annotation of the underlying genes are summarized. SNPs are categorized into three groups: reproducible across sites, reproducible within a site and reproducible but no putative gene detected within 50 kb interval. Scores of -log10 P ≥ 6 are in bold, genes with SNPs within its genomic sequence are highlighted in yellow.

Supplementary Table 2 and 3 are in excel format submitted as digital files.
Chapter 46: QTL mapping of stomata density in hybrid *Populus*

**Abstract**

Stomatal development is known to be regulated by various genetic and environmental factors to optimize gas exchange. Stomata in the epidermal tissues of leaves are valves that passes CO₂, and thus they influence the global carbon cycle, and during leaf development, react to the fluctuations on temperature, humidity and other climate changes. Genotypic variation in stomatal density has been reported but little is known of the genetic mechanisms behind these leaf traits. With the genome sequenced for the model tree *Populus trichocarpa*, identifying quantitative trait loci (QTL) for stomatal density will pinpoint polymorphisms linked to genes and provide a foundation for future work. In the present study, a hybrid pseudo-backcross poplar family was used to map QTL for stomatal density across two different environments and three years. A total of eighteen QTL were detected in eleven linkage groups, which explained 4.6% to 15.6% of the phenotypic variance. Four major genomic regions, with reproducible and consistent QTL clusters associated with stomatal density, were found on linkage groups II, III, XVI and XIX. Further study is needed to fully understand the mechanism underlying the observed genetic association and to elucidate the function of the QTLs involved.

---

6 The material contained in this chapter is to be submitted to Tree Genetics and Genomics.
Introduction

In plants, stomata, bounded by a pair of guard cells, control gas exchange such as water vapor and CO$_2$ between the plant and its environment, providing an active contribution to photosynthesis and transpiration [1]. By sensing different exogenous signals, the stomatal complex balances CO$_2$ uptake versus water loss under continually changing environmental conditions [2, 3]. Breeders are continuously working to improve crop productivity and stress tolerance. Knowledge of the underlying molecular genetic basis of traits related to plant productivity and adaptation, such as stomatal development, can assist breeding efforts [4].

Fossil leaves records of the stomatal density (SD) spanning the past 400 Myr supported the predicted changes in atmospheric CO$_2$, as such provides key information in light of predicted changes in precipitation and temperature, as well as the CO$_2$ environment, during leaf development [3, 5, 6], which can impact biomass production [7]. SD varies among individual leaves, among individuals within and among populations and among plant species [8]. Moreover, the regulation of stomatal numbers is controlled by environmental signals including CO$_2$ concentration, light intensity, temperature and water deficit [9-11]. For instance, increased concentrations of atmospheric CO$_2$ result in the reduction of stomatal density [12-14]. So far, a handful of genes have been found to be involved in stomatal development [15-19]. Thus, the knowledge of signaling pathways by which environmental signals control stomatal development remains incomplete. Alternatively, some authors suggest that stomatal development is controlled by more complex scale-free networks [3, 20].

As scale-free networks are composed of many interconnected nodes, they are robust and exhibit a high degree of tolerance to node removal but are also fragile to hub removal [21]. Both cases are found in stomatal development for such its signaling is suggested to be organized as a scale-free network [3]. For example, recent data revealed that guard cells continue to develop regardless of the loss of certain signaling components, such as mutations in the blue light receptor genes resulting in stomata fail to respond to blue light [22, 23], as such exhibiting high tolerance to nodal removal. However, other essential components, including ABA, extracellular calcium ion, hydrogen peroxide, CO$_2$ and IAA signaling, can have a direct effect and cause a
significant increase in guard cell number [2, 24-29]. Thus, many genes and pathways are involved in stomatal development. Through QTL (Quantitative Trait Locus) mapping, the underlying genomic regions that contribute to complex traits such as stomatal number can be identified, and their effect quantified. With the availability of molecular markers, sequence data and genetic linkage maps, QTL analysis has been critical to uncover the genetic architecture of complex, adaptive traits.

The genus *Populus* L. has been adopted as a model system in forest genetics and tree physiology [30-32]. Due to the variation in adaptive traits within various poplar species, countless opportunities are available to study and isolate underlying genes associated with ecologically and/or economically important traits [32, 33]. QTL mapping has been applied to analyze the genetic variation underlying stomatal density for several species, including *Arabidopsis thaliana*, *Quercus robur*, *Oryza sativa* and *Vicia faba* [6, 34-36]. In this study, we apply the QTL mapping method to identify chromosomal regions controlling stomatal density in a hybrid pseudo-backcross poplar family for the purpose of identifying underlying candidate genes for major QTL associated with stomatal density.

**Materials and methods**

*Mapping population and plant material*

We utilized a hybrid poplar pedigree (Family 52-124) composed of 396 genotypes was generated by crossing the hybrid female clone 52-225 (TD), an F₁ hybrid derived from a cross between *P. trichocarpa* (TT, clone 93-968) and *P. deltoides* (DD, clone ILL-101), with *P. deltoides* (DD, clone D124) [37, 38].

One vegetatively propagated rooted cutting from each genotype of the mapping population was established in one-gallon pots in a greenhouse at Michigan Technological University in the fall of 2009 (Drost et al. 2015). Pots were moved outside the greenhouse in the fall of 2010. On June 2, 2011, 197 genotypes were planted at a field site (2 m x 2 m spacing) at the Ford Forestry Center of Michigan Technological University in Alberta village (Michigan, 46°38′37″N, 88°28′46″W). Two additional biological replicates of each genotype in the mapping pedigree were established between
February 22 and 26, 2016 in a greenhouse at Michigan Technological University, in one-gallon pots.

For stomata measurements, three leaves per plant grown in the greenhouse were analyzed in August 2010 from 90 genotypes. In 2016, leaves collected from two replicates of 150 genotypes grown in greenhouse were analyzed. Finally, in 2017, leaves collected from one replicate of 150 genotype grown in the greenhouse was analyzed.

**Stomatal imprints and density measurements**

Healthy leaves of the first seasonal flush were collected, pressed and dried. The leaves were covered with a thin layer of clear nail polish on the lower epidermis between the second and third vein, covering an oval spot, and then were left to dry. A leaf imprint was obtained with a strip of clear tape and the imprint was transferred to a microscope slide. Stomatal density replicas were counted on microphotographs obtained with the aid of a light confocal microscope at 20X magnification from a 10 µm in diameter sections of the leaf imprint.

To assess correlations among replicates and between years, pair-wise phenotypic correlations were calculated as Pearson correlation coefficients for stomatal density across the three years using WinSTAT [39] (Table S1).

**Genetic linkage map**

We used the genetic map of the mapping family 52-124 comprised of 3,568 SNP markers with known genomic positions for QTL identification. SNP genotyping, marker curation and genetic map construction were previously described by Muchero et al. 2015 and Bdeir et al. 2017 [37, 40].

**QTL analysis**

The data had no outliers, recording errors were corrected or deleted. Each dataset was checked for normal distribution using WinSTAT [39]. Transformations were deemed unnecessary (Figure S1). The data were analyzed for the presence of QTL using the MapQTL6 software [41]. To map QTL intervals on the genetic linkage map and to test for reproducibility across years and environments, the interval mapping method was
used. Furthermore, composite interval mapping with the Multiple-QTL Model (MQM) method was applied to further refine the QTL regions. The putative QTL were subjected to 1,000 genome-wide (GW) and chromosome-wide (CW) permutations to determine LOD significance thresholds at the 0.05 significance level [42]. These parameters were used for declaring the existence of a significant QTL.

**Candidate genes**

Since QTL regions are genome-anchored, the underlying genes in the intervals were identified from the *Populus* genome assembly V3.0 [43] in the Phytozome database using the BioMart tool (https://phytozome.jgi.doe.gov). Because of large QTL intervals on some of the linkage groups, underlying genes within intervals defined by MQM mapping were identified for the eighteen QTL. Intervals spanning the genomic regions (physical location by MQM mapping) summarized in Table 1 were used to identify the putative candidate genes. The Gene Ontology (GO) of the genes underlying the LOD maxima were found using the *A. thaliana* gene annotation and AgriGO with p<0.05 [44].

**Results**

*Repeatability of stomatal density measurements*

Correlations among replicates and years are shown in Table S1. Stomatal density showed a high correlation between replicates and across sites and years. The phenotypic correlations between replicates within the same year are as high as r=0.98 for different leaves of the same plant in samples of 2010 and around r=0.96 for clonal replicates in 2016 (p≤0.0001 for all comparisons). High correlations were also observed for comparisons among years (r=0.81 to 0.96, p≤0.0001).

Frequency distributions of stomatal densities in all years are shown in Figure S1. All datasets follow the pattern of normally distributed data.

*QTL for stomatal density*

Twenty-four individual QTL were detected for stomatal density on twelve linkage groups (Table 1), of which eight had LOD scores above the GW threshold. For all datasets, QTL were anchored to the *Populus* genome assembly (Fig. 1, Table 1). For QTL
found on linkage group II, VI, VII, XII, XV, XVIII and XIX, the P. deltoides genotype DD was associated with the higher value for stomatal density, while for QTL on linkage group III, VIII, X, XIII, XIV and XVI, the DT genotype was associated with the higher value of stomatal density (Table 1). The percentage of phenotypic variance explained (PVE) by individual QTL ranged from 4.6 to 15.6%. Figure 2 shows a graphical outline of the LOD score profiles for all individual QTL mapped on linkage groups after cofactor selection. Specific genomic regions with significant QTL including LG II, III, VIII, XII, XIV, XVI and XIX also displayed a suggestive QTL with low LOD score not reaching the CW threshold representing different replicates/years.

Based on reproducibility across replicates and years, four QTL clusters stood out on LG II, III, XVI and XIX representing thirteen individual QTL that were reproducible across at least two replicates and/or years with overlapping intervals and LOD maxima within proximity (Fig. 2, Table 1). In the first cluster on LG II, two QTL with LOD maxima between 154-173 cM were detected for one replicate in 2010 and for one clonal replicate in 2016. These QTL explain a comparatively high percentage of the total phenotypic variance (4.6% and 14.5%, respectively) and one of these two QTL was significant above the GW threshold. A suggestive peak was also detected in 2017. Five QTL found on LG III with LOD maxima within an interval between 75-97 cM were detected across all three years, the two clonal replicates in 2016, and in one replicate and their average in 2010, while a suggestive QTL below the CW threshold (LOD score=1.8) was detected between 73-95 cM for one replicate (leaf a) in 2010 (Fig. 2). The phenotypic variance explained by these QTL ranged between 5.5 and 12.7% and the QTL for replicates a in 2016 and replicate m in 2010 were significant at the GW level. The third QTL cluster consisting of three QTL were detected on LG XVI in 2010 and 2016 with LOD maxima at 20 cM. For 2016, one replicate (leaf a) showed a QTL significant at the CW threshold (PVE 6.2%), while for 2010, all replicates revealed a peak within same genomic regions. Specifically, replicates a and m had QTL significant at the GW threshold (PVE 12.3% and 11%) while the other two replicates of year 2010 (leaf b and c) showed a suggestive QTL within the same interval. Lastly, in the fourth QTL cluster on LG XIX an overlapping QTL interval was found between 0-40 cM for two replicates (leaf a, c and m) in 2010, while the third replicate (leaf b) showed a suggestive QTL.
within the same interval (Fig. 2). Though only one replicate was significant at the GW threshold, both QTL explain a high percentage of the total phenotypic variance (11.0% and 15.6%).

On LG XIII and XVIII, they both showed reproducibility across 2010 replicates, but the four QTL were significant at the CW threshold. The remaining 7 individual QTL were significant but not reproducible across replicates or years (Fig. 2, Table 1). QTL detected in year 2016 on LG VIII and another QTL on LG XIV, while both had high LOD score and PVE values (5.3 and 8.9%, respectively), their LOD maxima were at different positions (around 140 cM and 1 cM, respectively, Table 1). They both also showed one suggestive QTL within their interval across different replicates/years (Fig. 2). The third significant QTL (11SD10b, Table 1) on LG XII having high LOD and PVE scores (4.52 and 15.2%, respectively) overlapped with three suggestive QTL from replicates in years 2010, 2016 and 2017 (Fig. 2).

**Candidate genes**

The interval after MQM mapping with cofactor selection was used to identify putative genes and the number of underlying each QTL is listed in Table 2. A total of 869 genes within genome-anchored QTL intervals for stomatal density were detected, out of which 617 (71%) had annotations based on the InterPro domain; the function of the remaining genes is unknown. A detailed list of the putative genes with their annotations is given in Table S2.

**Significantly overexpressed GO terms**

To enable a functional categorization and to identify biological pathways of the genes detected, Gene Ontology (GO) term enrichment analyses were conducted. Figure S2 is a visual representation of all the underlying genes found within the eighteen QTL intervals and of significantly overrepresented GO terms in biological process categories. Figure S3 lists the significant levels of the represented GO terms.
Discussion

With the aid of QTL mapping, we can pinpoint the genomic regions associated with a quantitative trait and the polymorphisms that are linked to functional genes. Several reviews stated the advantages of using QTL analysis to reveal the genetic architecture especially for complex traits [45-47]. With the availability of the Populus genome, candidate genes in QTL intervals can be identified. In this study, we analyzed stomatal density in a hybrid Populus progeny using QTL analysis and identified major QTL responsible for stomatal density variation. It has been generally reported that stomatal density is controlled by both environmental and genetic factors, suggesting the involvement of many genes [48]. The signaling pathways and mechanisms that regulate guard cell function have been intensively studied (reviewed in [49, 50]). Furthermore, environmental variables can modulate basal stomatal development pathways to adjust to the environmental conditions [48]. For example, the concentration of atmospheric CO2 directly influences stomatal development and has been used an indicator of paleoatmospheric CO2 levels and in the assessment of the ecological consequences of global change [3, 12-14]. Forty percent of the atmospheric CO2 passes through stomata [51], thus any variation in stomatal density ultimately influences photosynthesis and atmospheric CO2 concentration. Our results however, based on strong inter annual correlations among genotypes and the correlations among replicates within years, suggest that genetics plays a strong role in determining stomatal density in hybrid Populus.

It has been suggested that guard cell signaling reflects the organization of a scale-free network rather than a collection of linear pathways [3]. The estimated number and effect of genes involved in such a complex trait most likely linked to a scale-free network can be obtained through QTL mapping. As a quantitative trait, it was genetically analyzed using QTL mapping in both poplar and oaks [6, 52]. Estimates for the effect of genes controlling stomatal density have been obtained in the QTL study by Rae et al. (2006) using a third-generation hybrid population derived from two Populus trichocarpa and P. deltoides hybrids. Specifically, the hybridization of P. trichocarpa (clone 93-968, same clone used in our study) and P. deltoides (clone ILL-129) generated two hybrids, 53-246 and 53-242, which were crossed to generate an F2 full-sib family (family 331). Rae et al (2006) performed QTL analyses under elevated and ambient CO2 conditions
and found fourteen QTL for adaxial and abaxial stomatal density distributed over ten linkage groups with genetically determined total phenotypic variances for elevated and ambient CO₂ conditions between 19.5-22.1% and 9.3-19.7%, respectively.

In another study, using a hybrid poplar family of Populus trichocarpa × P. deltoides saplings, stomatal development was monitored under varied CO₂ concentrations, vapor pressure deficit (D), and irradiance (Q) around the young leaves [53]. Results indicated that changes in CO₂ concentration, light and humidity causing an increase in mature leaf stomatal conductance had a regulatory effect on stomatal development of expanding leaves [53]. A plastic response to changing environments has been shown in stomatal density in maize, Arabidopsis and poplar for which systemic signals from mature leaves regulate stomatal development of expanding leaves [20, 53, 54]. However, the nature of these signals generated in the mature leaves and transmitted to developing leaves is largely unknown and it is unclear how many signals are elicited [55]. Our study further suggests high genetic variation for and control of stomatal density in a single interspecific hybrid progeny, but also variation among years and replicates.

QTL for stomatal density

We detected twelve genomic regions on eleven linkage groups with a range of QTL effects explained from 4.6% to 15.6 % of the phenotypic variance. Four major QTL clusters with high reproducibility and consistency associated with stomatal density were found on linkage group II, III and consistently identified across all three years, while the clusters on linkage group XVI were found reproducible across two years. Finally, QTL clusters on linkage group XIX were consistently identified across all three replicates in year 2010. Since the identified genomic intervals are large and contain hundreds of candidate genes, it is still difficult to identify the genes in question. Thus, we seek to analyze the overrepresented GO terms in the QTL intervals.

GO analysis showed the functional profile of the underlying genes and their involvement in biological processes such as developmental processes, including cell morphogenesis and differentiation and post-embryonic development, and cellular processes, such as metabolic and cellular biosynthesis regulations (Figure S2). It has been demonstrated that mature leaves both detect CO₂ concentration levels and accordingly
signal to expanding leaves thus inducing stomatal development in young leaves [13, 20, 56]. Furthermore, numerous studies showed a relationship between stomatal traits (density, index and conductance) with leaf development (expansion, length and area) [38, 52, 57, 58].

**Future Work**

The natural variation of stomatal density is linked to plant fitness and adaptation to their environment, thus directly impacting speciation and evolutionary change [3]. The availability of the full annotated poplar sequence and associated genomic resources including microarrays will be essential to further study and enable research to be focused on understanding stomatal development and the genetic control of this trait. Future work on nucleotide variation in candidate genes controlling such complex adaptive traits as stomatal density under different environmental conditions such as water deficit and different CO₂ concentrations will allow a better understanding of the genetic and environmental control of this trait.
References


Figure 1 QTL anchored to the genome of *Populus trichocarpa* (V3 assembly, Kelleher et al. 2007). The actual map has a high marker density (average marker spacing: 5 markers per 4 cM). For illustration purposes, for each linkage group an evenly spaced selection of scaffolds is shown (one marker per 20 cM). The yellow regions on LGs represent QTL intervals using interval mapping (before co-factor selection). QTL for stomatal density (SD) are shown in red, green and blue for datasets collected in 2010, 2016 and 2017, respectively. The outer lines of bars are CW thresholds and the middle lines are LOD maxima (see Table 1). Scaffold intervals are represented in Mb.
Figure 2 LOD score profiles for stomatal density QTL on LG II, III, VI, VIII, X, XII, XIII, XIV, XV, XVI and XIX using the Multiple-QTL Model (MQM) with co-factor selection across all experimental replicates. Chromosome-wide (CW) and genome-wide (GW) significance thresholds are shown with dashed lines ($\alpha = 0.05$, 1,000 permutations). Profiles for stomatal density are shown in shades of red, blue and green for data sets collected in 2010, 2016 and 2017, respectively. Some suggestive QTL with lower LOD score than the CW threshold are shown as well.
Table 1. QTL associated with stomatal density identified in the poplar pseudo-backcross pedigree 52-124 for years 2010, 2016 and 2017.

<table>
<thead>
<tr>
<th>QTL #, population (lg)</th>
<th>Interval Mapping</th>
<th>MQM Mapping</th>
<th>LOD max</th>
<th>Origin of positive allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>Map location (V2, cM)</td>
<td>Physical location (V3)</td>
<td>Physical location (V3)</td>
<td>LOD value</td>
<td>Location (cM)</td>
</tr>
<tr>
<td>1SD16a (II)</td>
<td>6.321-7.943</td>
<td>2390513..2532397</td>
<td>2390513..2532397</td>
<td>2.34</td>
</tr>
<tr>
<td>2SD16b (II)</td>
<td>113.997-155.482</td>
<td>20910935..23483297</td>
<td>20910935..23483297</td>
<td>4.35*</td>
</tr>
<tr>
<td>3SD10a (II)</td>
<td>156.015-175.197</td>
<td>11654461..11956032</td>
<td>12307868..12635780</td>
<td>2.23</td>
</tr>
<tr>
<td>4SD16b (III)</td>
<td>74.564-89.295</td>
<td>2390513..2532397</td>
<td>2390513..2532397</td>
<td>2.34</td>
</tr>
<tr>
<td>5SD16a (III)</td>
<td>156.015-175.197</td>
<td>20910935..23483297</td>
<td>20910935..23483297</td>
<td>4.35*</td>
</tr>
<tr>
<td>6SD10c (III)</td>
<td>82.228-100.001</td>
<td>2390513..2532397</td>
<td>2390513..2532397</td>
<td>2.34</td>
</tr>
<tr>
<td>7SD10m (III)</td>
<td>124.933-139.844</td>
<td>12307868..12635780</td>
<td>12307868..12635780</td>
<td>2.23</td>
</tr>
<tr>
<td>8SD17a (III)</td>
<td>104.001-116.113</td>
<td>11654461..11956032</td>
<td>12307868..12635780</td>
<td>4.01*</td>
</tr>
<tr>
<td>9SD10b (VI)</td>
<td>6.312-8.934</td>
<td>2390513..2532397</td>
<td>2390513..2532397</td>
<td>2.34</td>
</tr>
<tr>
<td>10SD16a (VIII)</td>
<td>137.888-141.713</td>
<td>11654461..11956032</td>
<td>12307868..12635780</td>
<td>4.01*</td>
</tr>
<tr>
<td>11SD16b (X)</td>
<td>75.586-92.973</td>
<td>11654461..11956032</td>
<td>12307868..12635780</td>
<td>4.01*</td>
</tr>
<tr>
<td>12SD10b (XII)</td>
<td>0.877-5.3703</td>
<td>11654461..11956032</td>
<td>12307868..12635780</td>
<td>4.01*</td>
</tr>
<tr>
<td>13SD16b (XIII)</td>
<td>6.312-8.934</td>
<td>2390513..2532397</td>
<td>2390513..2532397</td>
<td>2.34</td>
</tr>
<tr>
<td>14SD16b (XIII)</td>
<td>6.312-8.934</td>
<td>2390513..2532397</td>
<td>2390513..2532397</td>
<td>2.34</td>
</tr>
<tr>
<td>15D16b (XIV)</td>
<td>0.16.565</td>
<td>0.1979129</td>
<td>0.68230</td>
<td>4.17*</td>
</tr>
<tr>
<td>16SD16b (XV)</td>
<td>30.834-36.579</td>
<td>2608259..3099304</td>
<td>2608259..3099304</td>
<td>2.58</td>
</tr>
<tr>
<td>17SD16a (XVI)</td>
<td>19.291-24.68</td>
<td>1768332..2141944</td>
<td>1768332..1955790</td>
<td>2.41</td>
</tr>
<tr>
<td>18SD10a (XVI)</td>
<td>19.291-39.166</td>
<td>1588548..3292119</td>
<td>1588548..3292119</td>
<td>3.76*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------------</td>
<td>-----------</td>
</tr>
<tr>
<td>19SD10m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(XVI)</td>
<td>21.477-49.048</td>
<td>1975247..4163931</td>
<td>3582922..3913475</td>
<td>4.01*</td>
</tr>
<tr>
<td>20SD10a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(XVIII)</td>
<td>108.468-119.783</td>
<td>15087827..16199806</td>
<td>15910227..16065443</td>
<td>2.17</td>
</tr>
<tr>
<td>21SD10m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(XVIII)</td>
<td>112.31-119.611</td>
<td>15301185..16216454</td>
<td>15910227..16065443</td>
<td>1.97</td>
</tr>
<tr>
<td>22SD10c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(XIX)</td>
<td>0-33.215</td>
<td>0..8436001</td>
<td>0..389959</td>
<td>2.7</td>
</tr>
<tr>
<td>23SD10a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(XIX)</td>
<td>21.904-40.875</td>
<td>4152016..10085840</td>
<td>4554508..4983998</td>
<td>4.65*</td>
</tr>
<tr>
<td>24SD10m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(XIX)</td>
<td>0-40.875</td>
<td>0..10085840</td>
<td>4152016..4870964</td>
<td>4.45*</td>
</tr>
</tbody>
</table>

lg: linkage group; V2: markers anchored on version 2 of the *P. trichocarpa* genome; V3: version 3 updated physical location; PVE: percent phenotypic variance explained; DD: homozygous for the *P. deltoides* allele, DT: heterozygous for the *P. deltoides* and *P. trichocarpa* alleles. LOD max determined using MQM mapping, value with *: above GW threshold, otherwise above CW threshold (Example: 1SD16a: QTL number one, stomata density, year 2016, replicate a). The number of genotypes is 90 for year 2010 and 150 for both 2016 and 2017.
Table 2 Number of candidate genes detected across the twenty QTL significantly associated with stomatal density.

<table>
<thead>
<tr>
<th>QTL #, dataset (lg)</th>
<th>Interval (V3, bp)</th>
<th>Number of genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SD16a (II)</td>
<td>2390513..2532397</td>
<td>22</td>
</tr>
<tr>
<td>2SD16b (II)</td>
<td>16904784..18010025</td>
<td>45</td>
</tr>
<tr>
<td>3SD10a (II)</td>
<td>20910935..23483297</td>
<td>114</td>
</tr>
<tr>
<td>4SD16b (III)</td>
<td>11654461..11956032</td>
<td>34</td>
</tr>
<tr>
<td>5SD16a (III)</td>
<td>12307868..12635780</td>
<td>44</td>
</tr>
<tr>
<td>6SD10c (III)</td>
<td>13706345..13918404</td>
<td>27</td>
</tr>
<tr>
<td>7SD10m (III)</td>
<td>13457418..13606110</td>
<td>14</td>
</tr>
<tr>
<td>8SD17a (III)</td>
<td>14830986..15157363</td>
<td>42</td>
</tr>
<tr>
<td>9SD10b (VI)</td>
<td>7010101..7418569</td>
<td>52</td>
</tr>
<tr>
<td>10SD16a (VIII)</td>
<td>11733076..12070763</td>
<td>50</td>
</tr>
<tr>
<td>11SD16b (X)</td>
<td>12121301..12589424</td>
<td>66</td>
</tr>
<tr>
<td>12SD10b (XII)</td>
<td>3037452..3341456</td>
<td>35</td>
</tr>
<tr>
<td>13SD10b (XIII)</td>
<td>6076968..6442032</td>
<td>28</td>
</tr>
<tr>
<td>14SD10m (XIII)</td>
<td>6076968..6442032</td>
<td>28</td>
</tr>
<tr>
<td>15D16b (XIV)</td>
<td>0..68230</td>
<td>6</td>
</tr>
<tr>
<td>16SD16b (XV)</td>
<td>2608259..3099304</td>
<td>34</td>
</tr>
<tr>
<td>17SD16a (XVI)</td>
<td>1768332..1955790</td>
<td>24</td>
</tr>
<tr>
<td>18SD10a (XVI)</td>
<td>1588548..2141944</td>
<td>82</td>
</tr>
<tr>
<td>19SD10m (XVI)</td>
<td>3582922..3913475</td>
<td>34</td>
</tr>
<tr>
<td>20SD10a (XVIII)</td>
<td>15910227..16065443</td>
<td>15</td>
</tr>
<tr>
<td>21SD10m (XVIII)</td>
<td>15910227..16065443</td>
<td>15</td>
</tr>
<tr>
<td>22SD10c (XIX)</td>
<td>0..389959</td>
<td>38</td>
</tr>
<tr>
<td>23SD10a (XIX)</td>
<td>4554508..4983998</td>
<td>30</td>
</tr>
<tr>
<td>24SD10m (XIX)</td>
<td>4152016..4870964</td>
<td>54</td>
</tr>
</tbody>
</table>

Number of genes found within the nineteen QTL based on MQM mapping with cofactor selection. lg: linkage group.
Supplementary Figure 1 Frequency distribution for stomatal density collected in year 2010, 2016 and 2017 (a, b and c, respectively). Three leaves per genotype were tested for stomatal density in 2010, while leaves taken from two individual plants per genotype were sampled in 2016.
Supplementary Figure 2 Hierarchical tree graph of overrepresented GO terms in biological process categories for all underlying genes within the twenty trait-associated QTL. The non-significant terms are shown as white boxes, and significant terms ($P \leq 0.05$) are marked with color (the degree of color saturation of boxes is associated to the enrichment level of the term). Red indicates the highest enrichment level.

Supplementary Figure 3 The significant levels of the overrepresented GO terms colored from shades of yellow, orange and red as well as the relationships between GO terms indicated by different arrows are shown. Red indicates the highest enrichment level.
**Supplementary Table 1** Pair-wise estimates of phenotypic correlations calculated as Pearson correlation coefficients for stomatal density collected for years 2010, 2016 and 2017.

<table>
<thead>
<tr>
<th></th>
<th>SD10a</th>
<th>SD10b</th>
<th>SD10c</th>
<th>SD10m</th>
<th>SD16a</th>
<th>SD16b</th>
<th>SD17a</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD10a</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD10b</td>
<td>0.985032</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD10c</td>
<td>0.988059</td>
<td>0.981475</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD10m</td>
<td>0.976836</td>
<td>0.98128</td>
<td>0.985206</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD16a</td>
<td>0.92854</td>
<td>0.926599</td>
<td>0.944615</td>
<td>0.960826</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD16b</td>
<td>0.949868</td>
<td>0.9464</td>
<td>0.967087</td>
<td>0.966766</td>
<td>0.980137</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>SD17a</td>
<td>0.91097</td>
<td>0.898656</td>
<td>0.924961</td>
<td>0.930541</td>
<td>0.939645</td>
<td>0.961409</td>
<td>X</td>
</tr>
</tbody>
</table>

The analysis for stomatal density (SD) was done between replicates and across years. All correlations had a p-value < 0.0001. For 2010, replicates a, b and c represent leaves taken from the same plant and replicate m is their mean; whereas for 2016, replicates a and b are clonal replicates taken from two individual plants per genotype.

**Supplementary Table 2** All candidate genes within the twenty QTL detected for stomatal density variation in *Populus*. Physical localization and annotation of gene models within each QTL interval are listed.

---

7 Supplementary Table 2 are in excel format submitted as digital files.
Chapter 5: Summary and Conclusion

The present thesis investigates the genus *Populus* using population genomics approaches to better understand the genetic system controlling the inheritance of bark features, diameter growth and stomatal density and to associate these traits with respective genes.

For Quantitative Trait Locus (QTL) analysis, the experimental material consisted of an interspecific hybrid poplar pseudo-backcross pedigree composed of 396 genotypes. The mapping population was planted at Oregon (OR) and West Virginia (WV) in a three-block replication for a total of six ramets per cloned genotype. Overall, bark texture, bark thickness and diameter growth showed high correlations among replicates and years, however, across sites, correlations were only significant for bark texture.

A total of 94 individual QTL detected for the three traits, bark texture, bark thickness and stem diameter, across various chromosomes were successfully anchored to the *Populus* genome assembly. Specifically, five QTL clusters for diameter and seven QTL clusters for bark thickness were observed. The most significant and reproducible results were found associated with bark texture for which seven QTL clusters were detected. Given the environmental contrast between the OR and WV experimental sites, four out of the seven QTL clusters, representing a total of 47 individual QTL detected for bark texture, were remarkably consistent across both sites. Differences in reproducibility for QTL clusters across sites suggest differential environmental effects on gene expression. In comparison to bark texture, QTL clusters for bark thickness and diameter had lower reproducibility across sites.

QTL clusters on chromosome I, VI and XII were associated with all three traits. Co-location of QTL for traits can be the result of pleiotropic effects or closely linked genes. These overlapping QTL could be an explanation of different aspects of bark texture, bark thickness and radial growth. Romero [1] proposed that rough bark results in response to the mechanical stresses imposed by a varied radial growth and due to different meristematic activity in the phellogen, a discontinuous periderm. Strong correlations between bark texture and diameter could indicate that bark texture is partly related to diameter growth. Furthermore, using MQM mapping, QTL for these traits were mapped to different neighboring positions of the same chromosomes (Fig. 2).
Consequently, bark texture seems to be only partly related to diameter growth, and other factors such as meristematic activity of the phellogen and cell adhesion are likely to have major effects on bark texture. A higher mapping resolution as obtained in linkage disequilibrium mapping in natural population samples is needed to narrow down QTL regions to individual genes and to distinguish between pleiotropic effects and close linkage.

This limitation was evident in our QTL analyses while looking at the candidate genes list, where most clusters included from 123 to 963 candidate genes. Nonetheless, several candidate genes within the QTL interval can be identified based on their putative functions including PopNAC128 (Potri.001G206900), one of the orthologs of Arabidopsis ANAC104 (Arabidopsis Nac Domain Containing Protein 104) and XND1 (Xylem NAC Domain 1), and Potri.001G206700, an ortholog of AT4G33430 (BAK1, Br1-Associated Receptor Kinase; ELONGATED; SERK3). The first gene was shown to be involved in the development of phloem fibers and its overexpression resulted in a slowed secondary phloem development [2], while the second gene is found to be involved in patterning and growth regulation [3, 4].

Variation in bark texture could be related to cell adhesion which is essential to form a single periderm resulting in smooth bark. The lack of cell-cell adhesion leads to the development of uneven and discontinued bark or bark splitting causes a peeling and fissured bark appearance. At the molecular level, several expressional candidate genes identified in this study have a role in cell adhesion, including Pinin (Potri.001G208200) and PopFLA or Fasciclin-Like Arabinogalactan (Potri.013G151300, Potri.013G151400 and Potri.013G151500)[5-7]. They all fell within QTL intervals with the highest LOD scores and had high expression in phloem and xylem tissues and the cambium.

Bark features in our study ranged from smooth to deeply furrowed which is characteristic for P. deltoides. Variation in shallowly fissured bark which is characteristic for P. trichocarpa was not observed in the segregating progeny. Thus, the QTL identified in this progeny set only represent a subset of a larger number of polymorphisms affecting the traits. And in our pseudo-backcross pedigree (DD X DT) involving multiple P. deltoides alleles, polymorphisms associated with characteristic bark features of P. trichocarpa seem to be largely undetected. Association populations for P. trichocarpa
will be used to find additional candidate genes associated with bark texture in this species.

Therefore, a Genome-Wide Association Study (GWAS) analysis was carried out using an association population consisting of 1,100 black cottonwood genotypes \((Populus trichocarpa)\) Torr. & Gray). They were planted at Oregon (OR) and California (CA) where genotypes were planted in a three-block replication for a total of six ramets per genotype. Our results revealed multiple significant associations. Through GWAS, we were able to detect several putative genes involved in the control of bark texture, as well as differentiate between the genes responsible for the flaky bark texture in \(P. trichocarpa\) and the rough furrowed texture in \(P. deltoides\) in comparison to the QTL study.

Overall, the detected genomic regions involved in the control of bark texture in \(P. trichocarpa\) and \(P. deltoides\) vary, only eight genes out of 380 GWAS-identified genes were found to be co-locating with the QTL study.

A candidate gene list based on the highest p-values and reproducibility with possible functional control of bark texture was identified. Their putative function fell within three different categories: 1) radial growth and tissue differentiation, 2) suberin accumulation and 3) programmed cell death. Vascular differentiation and differentiation of the outer bark involves several steps including cell division, orientated cell differentiation, cell expansion, cell wall thickening, and programmed cell death. Discontinuous periderms (textured bark) could result from variable radial meristematic activity in the cork cambium due to mechanical stresses from radial growth. Phellem is multilayered dead tissue that is made impervious by the disposition of suberin onto their cell walls while suberization occurs during the development of the secondary radial meristems (the cork cambium). Finally, programmed cell death is an essential component during the plant’s secondary development and is involved in the generation of the vascular system contributing to the development of bark and its abscission.

All three processes are essential to secondary development and can influence the development of bark. Thus, a better understanding of the putative genes involved in these developments can provide insights into bark tissue development. Several genes of unknown function detected within highly significant regions associated with bark texture open new opportunities for future studies and should be functionally characterized.
The combination of QTL and GWAS mapping supports the identification of candidate genes. Eight genes out of 380 GWAS-identified genes were found to be co-localizing with the QTL from the previous study, five of which have an unknown function. Two genes with annotated function as root hair defective 3, Potri.012G116900 and Potri.012G117000, co-located with major highly reproductive QTL for bark texture. The same QTL region was also associated with bark thickness and diameter, suggesting a role of this QTL in radial growth. These genes putatively function as transmembrane proteins identified from the plasma membrane of *Populus* differentiating xylem and phloem [8].

QTL mapping increases the ability to reduce false positive rates whereas GWAS has a much higher resolution than QTL mapping and can narrow down the genomic region to individual candidate genes. In our case, the combination of QTL and GWAS mapping provided insights into the different genes responsible for variations in bark texture ranging from smooth and furrowed across *Populus* species.

Looking at the stomatal density results using QTL mapping, four QTL clusters were detected along with several individual QTL. It has been generally reported that stomatal density is controlled by both environmental and genetic factors, suggesting the involvement of many genes. Any variation in stomatal density ultimately influences photosynthesis and atmospheric CO₂ concentration. Our results however, based on strong inter annual correlations among genotypes and the correlations among replicates within years, suggest that genetics plays a strong role in determining stomatal density in *Populus*.

Systemic signals from mature leaves regulate stomatal development of expanding leaves, however, the nature of these signals generated in the mature leaves and transmitted to developing leaves is largely unknown and it is unclear how many signals are elicited. Our study further suggests high genetic variation for the control of stomatal density in a single interspecific hybrid progeny, but also variation among years and replicates. We detected twelve genomic regions on eleven linkage groups with a range of QTL effects explaining from 4.6% to 15.6 % of the phenotypic variance. Four major QTL clusters with high reproducibility and consistency associated with stomatal density were found on linkage groups (LGs) II, III, XVI and XIX. Clusters on LG II, III were consistently identified across all three years, while the clusters on LG XVI were found
reproducible across two years. Since the identified genomic intervals are large and contain hundreds of candidate genes, it is still difficult to identify the genes in question. Thus, we seek to analyze the overrepresented GO terms in the QTL intervals. GO analysis showed the functional profile of the underlying genes and their involvement in biological processes such as developmental processes, including cell morphogenesis and differentiation, post-embryonic development, and cellular processes, such as metabolic and cellular biosynthesis regulations. The natural variation of stomatal density is linked to plant fitness and adaptation to their environment, thus directly impacting speciation and evolutionary change [9].

QTL and GWAS mapping provide a baseline, a list of candidate genes associated with the trait of interest. Future work through transgenic lines and assessment of nucleotide variation in these candidate genes controlling such complex adaptive traits under different environmental conditions will allow a better understanding of the genetic and environmental control of these traits.