Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2021:60
Faculty of Natural Resources and Agricultural Sciences

Opportunities in small population breeding in black cottonwood

or

The little population that could

Rami-Petteri Apuli

Opportunities in small population breeding in black cottonwood

or
The little population that could

Rami-Petteri Apuli
Faculty of Natural Resources and Agricultural Sciences
Department of Plant Biology
Uppsala

SLU
SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES

DOCTORAL THESIS
Uppsala 2021

Acta Universitatis Agriculturae Sueciae
2021:60

Cover: The Krusenberg CLAP black cottonwood in autumn colors.
(Photo: Thomas Richards)

ISSN 1652-6880
ISBN (print version) 978-91-7760-797-7
ISBN (electronic version) 978-91-7760-798-4

Uppsala
Print: SLU Service/Repro, Uppsala 2021

Opportunities in small population breeding in black cottonwood

Abstract

Small and structured populations are problematic for breeding due to low levels of genetic variability and increased levels of linkage disequilibrium (LD). The increased LD causes problems with utilization of common genomic tools and may lead into increased accruement of deleterious alleles through the intense selection and bottlenecks characteristic of breeding. In forest trees, maintaining a large breeding population to alleviate these issues is difficult due to large physical size of most trees. The aim of this thesis was to study and resolve some of the LD dependent issues of small populations and establish small populations as an option in breeding.

We established the accuracy of LD-based estimation methods of recombination using data from European aspen, allowing confident use of these methods later in the project. The small black cottonwood breeding population contained enough genetic diversity to facilitate future adaptive selection to novel Swedish climate and light conditions, suggesting that small populations consisting of offspring with diverse parentage are feasible options for breeding in species with high outcrossing and recombination rates. We identified candidate genes that can be targeted for selection on phenology and growth using a genome wide association study (GWAS). We also show that GWAS is useful for identifying large effect alleles even in small populations and that efficient growth under novel conditions likely require different allele combinations than in native habitats. Finally, we quantified deleterious load and identified effects of the deleterious load on growth. Accounting for deleterious load allowed for more effective genomic selection and increased breeding cycle gain in breeding programs based on small populations.

Keywords: Plant breeding, forest trees, small population, genetic linkage, linkage disequilibrium, GWAS, local adaptation, deleterious load, *Populus*

Author's address: Rami-Petteri Apuli, Swedish University of Agricultural Sciences, Department of Plant Biology, Uppsala, Sweden

Pienpopulaatiojalostamisen mahdollisuudet Jättipoppelissa

Lyhennelmä

Pienet populaatiot ovat ongelmallisia jalostustoiminnassa johtuen niiden vähäisestä geneettisestä monimuotoisuudesta ja lisääntyneestä kytkentäepätasapainosta (LD). Korkea LD-taso aiheuttaa ongelmia useiden yleisten genomisten työkalujen käyttössä ja voi johtaa geneettisen kuorman runsastumiseen jalostustoiminnalle tunnusomaisten valinnan ja pullonkaula-ilmiöiden kautta. Kohonnut LD on eritoten ongelmallinen metsäpuille, joiden suuri koko tekee tarpeeksi suurten jalostuspopulaatioiden ylläpitämisen haasteelliseksi. Tämän väitöskirjan tavoitteena oli tutkia ja ratkaista osa näistä LD:n aiheuttamista ongelmista ja osoittaa pienten populaatioiden olevan varteenotettava vaihtoehto jalostustoiminnassa.

Varmistimme LD-rekombinaatioarviointimenetelmien tarkkuuden käyttäen metsähaapadataa. Jättipoppelipopulaatiomme on pienestä koostaan huolimatta tarpeeksi geneettisesti monimuotoinen mahdollistamaan lajin sopeuttamisen Ruotsin ilmasto- ja valo-oloihin. Osoitimme pienten jälkeläispopulaatioiden olevan soveltuvia jalostukseen mikäli ne on tuotettu monimuotoisista vanhemmista. Löysimme useita ehdokasgeenejä ilmastosopeuttamisen valintakohteiksi käyttäen genominlaajuista assosiaatiotutkimusta (GWAS) osoittaen samalla GWAS:n olevan käytännöllinen tapa havaita suurivaikutuksisia alleeleja myös pienissä populaatioissa. Totesimme myös kasvun uusissa oloissa vaatinevan uusia alleeliyhdistelmiä kuin lajin alkuperäisoloissa. Lopuksi havaitsimme geneettisen kuorman vaikuttavan kasvuun, joten sen huomioiminen jalostettaessa mahdollistanee tehokkaamman genomisen valinnan pienissä populaatioissa.

Avainsanat: Kasvinjalostus, metsäpuut, pienpopulaatio, geneettinen kytkentä, kytkentäepätasapaino, GWAS, paikallinen sopeutuminen, geneettinen kuorma, *Populus*

Kirjoittajan osoite: Rami-Petteri Apuli, Swedish University of Agricultural Sciences, Department of Plant Biology, Uppsala, Sweden

Preface

Well, here we are. Not in a million years would I have believed that some 26 years after my little 5-year-old eyes gazed upon the Saturday night nature documentaries for the first time and loved them, would I be writing the preface of my very own thesis in the field of biology. It is even more amazing seeing that I never planned to go for PhD after attaining my master's degree, not necessarily because I did not want to, but because I doubted I could. I never saw in myself the qualities of a great scientist, and still more often do not than do, but as I said: here we are, no matter how unexpected, thanks to positive mindset, often rooted in the belief from others, and persistency.

Then again, during my time working on this project, the unexpected has become quite commonplace and the PhD project I began with is far from the PhD I ended with. Through one of these unexpected twists, I ended up working with the "CLAP population", a group of black cottonwood individuals with a mysterious past and flawed character for scientific purposes, yet somehow destined to potentially be the basis of a breeding program. I have often been at odds with this population and much like many of the reviewers for the papers constituting this thesis, I have often turned my nose in disgust at its flaws, wishing for a better, more diverse and abundant population to analyse. Over time I came to realize this population represents very much real life, where perfect scientific situations rarely exist. This flawed little population deserved to be studied as much as the theoretically perfect one to uncover real-life solutions to real-life problems.

And so, through hard work of myself and others, the flawed little population has proved its capabilities to fulfil its projected destiny and, on the side, proved that I could become a PhD and write a thesis. My thesis, thesis of the 5-year-old nature doc fan and thesis of the unexpected scientist, based on the unexpected little population that could. I hope you enjoy it.

Dedication

To the loving mother who took a little boy to see migrating birds in the spring.

To the caring father who took a little boy to fish in the summer.

To the little boy who never stopped being curious of the living world.

"Science is simply common sense at its best." – Thomas Henry Huxley

Contents

List of publications

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

I. Apuli, R.-P., Bernhardsson, C., Schiffthaler, B., Robinson, K.M., Jansson, S., Nathaniel R. Street, N.R., and Ingvarsson, P.K.. (2020). inferring the genomic landscape of recombination rate variation in European aspen (*Populus tremula*). *G3: Genes, Genomes, Genetics* 10 (1), pp. 299–309. https://doi.org/10.1534/g3.119.400504.

II. Richards, T.J., Karacic, A., Apuli, R.-P., Weih, M., Ingvarsson, P.K., and Rönnberg-Wästljung, A.-C.. (2020). Quantitative genetic architecture of adaptive phenology traits in the deciduous tree, *Populus trichocarpa* (Torr. and Gray). *Heredity* 125 (6) pp. 449-458. http://dx.doi.org/10.1038/s41437-020-00363-z

III. Apuli, R.-P., Richards, T.J., Rendón-Anaya, M., Karacic, A., Rönnberg-Wästljung, A.-C., and Ingvarsson, P.K.. (2021). The genetic basis of adaptation in phenology in an introduced population of Black Cottonwood (*Populus trichocarpa*, Torr. & Gray). *BMC Plant Biology*, 21(1), pp. 317. http://dx.doi.org/10.1186/s12870-021-03103-5

IV. Apuli, R.-P and Ingvarsson, P.K.. Deleterious load affects growth and breeding cycle gain in small population of Black Cottonwood (*Populus trichocarpa*, Torr. & Gray). Manuscript.

The contribution of Rami-Petteri Apuli to the papers included in this thesis was as follows:

I. Basic bioinformatics work, analyzed the data and took lead in writing the paper.

II. Set up part of study material, collected data and contributed to writing process of paper.

III. Set up part of study material, collected data, performed basic bioinformatics work, planned some of the analyses, analyzed data and took lead in writing the paper.

IV. Collected data, performed basic bioinformatic work, planned some of the analyses, analyzed data and took lead in writing the paper.

List of tables

List of figures

Abbreviations

bp	Basepair(s)
cM	CentiMorgan
DNA	Deoxyribonucleic acid
GEBV	Genomic breeding value
GWAS	Genome wide association study/studies
LD	Linkage disequilibrium
LDB	Linkage disequilibrium based
LMB	(Genetic) Linkage map based
LS	Lindley score
SNP	Single nucleotide polymorphism

1. Introduction

1.1 Basis and functions of genetic linkage

The phenomenon known as genetic linkage was first reported by William Bateson, Edith Rebecca Saunders and Reginald Punnett in 1905 (Bateson et al. 1905), some five years after the rediscovery of Gregor Mendel's iconic work first performed in 1865 (Benson 2001). Prior to this, each phenotypic trait was believed to be inherited separately from one another (Bateson et al. 1905, Benson 2001). It was only a few years later that the idea of physical connection between the genes seen to exhibit genetic linkage was proposed by Morgan (1910), leading to formulation of the first concepts of linkage in chromosomes and linkage groups, and the particulate form of inheritance, i.e. that genes are contained in physical entities (Morgan 1910; Benson 2001). It was Morgan's student, Alfred Sturtevant, who first calculated a linkage map for multiple genes on same linkage group (Morgan 1915), formulating the concept of the still commonly used unit of genetic distance Morgan (M), one of which roughly translates to a distance with one recombination event per meiosis. The more commonly used unit is centiMorgan (cM), which, with some considerations, translates to the likelihood of a recombination event occurring between two sites during meiosis.

As posited by Morgan (1910), genetic linkage results from the physical attachment of each nucleotide to the two adjacent nucleotides on the DNA molecule forming a chromosome. The chromosome these nucleotides form is thus akin to a chain, where every link is attached to all the others through the other links on the chain. In a theoretical situation, where the chain cannot be altered in any way, the chain must be selected, rejected or passed down to

offspring as a single object with no regard for the merits or misgivings of each individual link in the chain. In reality, changes occur in the genetic sequence of a chromosome through mutations and other phenomena such as transposable elements and gene conversion. We will focus on mutations as they are the most relevant to the research in this thesis. Mutations are changes in the genetic sequence arising usually from inaccurate repair of damage to the genetic sequence (Chen et al. 2014; Chen & Furano 2015). Mutations can be additions, removals or rearrangements of parts of the DNA sequence, or replacements of one or more nucleotides in the sequence (Carlin 2011). The most common type of mutations, single nucleotide changing mutations may lead to single nucleotide polymorphism (SNP), if they remain and segregate within a population. Mutations thus create novel variants of the chromosome with differences in the nucleotide sequence and as such are at the base of genetic diversity (Carlin 2011).

1.1.1 Recombination breaks genetic linkage

Mutations in genes, promoters and other functional elements of the genome are almost always harmful as they disrupt the functionality of the cell they occur in. As such, these disruptive or deleterious mutations, as they are often called, lead to lowered or altered functionality, or even death, in the cell that contains them (Haldane 1935, reprinted in Haldane 2004). In haploid (1*n*, where *n* is the number of unique chromosomes) single cell organisms such as archaea and bacteria as well as some eukaryotes, a mutation is always expressed and in drastic cases can even lead to the death of the individual (Gerstein 2013). Evolution of sexual reproduction and polyploidy, and through them recombination, has been the solution to the problems posed by mutations and linkage for the majority of higher eukaryote species (Felsenstein 1974; Gerstein 2013). Recombination occurs only in sexually reproducing species with ploidy level of at least 2*n*, i.e. diploid species (as opposed to haploid 1*n*). Furthermore, this diploidy often allows for damaged copies of a gene to exists within an individual as long as there still remains a functional copy on one of the matching chromosomes, thus allowing individuals and cells to survive the damaging effects of the mutation(s) (Felsenstein 1974; Gerstein 2013; Orr-Weaver 2015).

Sexual reproduction, paired with polyploidy, leads to shuffling of genetic content in two ways. The first way is the act of sexual reproduction itself, where a male gamete and a female gamete combine to produce a zygote, an

offspring cell, where half of the chromosomes originate from the mother and the other half from the father, that can then develop into a new individual. Gametes are special reproductive cells containing only half of the chromosomes of the species, which is 1*n* in case of diploid (2*n*) (Figure 1). They are produced through a special cell division called meiosis, where the second way of shuffling of chromosomal contents occurs. In meiosis, the chromosomes pair up with another copy of the same chromosome before being split into separate gametes (Figure 1). This pairing is why higher, even numbered ploidy levels, i.e. polyploids, such as 4*n*, 6*n* and 8*n*, are feasible as it ensures that half of all chromosomes end up in separate gametes. During this process, the paired chromosomes are anchored to one another through structures called chiasma, and resolution of these chiasma is what may lead into chromosomal cross-over. Depending on how the chiasma are excised, parts of the paired chromosomes may switch places, i.e. cross over to the other chromosome, recombining the genetic material on the two chromosomes with one another and producing new hybrid chromosomes, thus potentially breaking linkage by creating new allele combinations across the genome (Figure 1) (Wright 1931; Felsenstein 1974).

This chromosome content shuffling is called meiotic recombination (hereafter recombination) and the rate at which it occurs is not even across the genome of a species but rather has been shown to be correlated with certain genomic features such as gene density (Giraut et al. 2011; Wang et al. 2016), GC-content (Kim et al. 2007; Giraut et al. 2011), repeat density, cytosine methylation (Giraut et al. 2011; Rodgers-Melnick et al. 2015) and levels of neutral genetic diversity (Nachman 2002) However, the magnitudes and directions of these associations are still under debate. Furthermore, effective recombination rates across the genome vary considerably between species (Henderson 2012; Tiley & Burleigh 2015) or even between the populations (Slavov et al. 2012) of the same species, due to differences in, for instance, demographic history and effective population size. Understanding the associations between recombination rate and various genomic features is therefore important at the genus, species and even population level (Henderson 2012; Slavov et al. 2012; Tiley & Burleigh 2015) to avoid making assumptions about the strength and direction of these associations (Paper I). Local variation in recombination rates is thus an important factor for understanding how natural or artificial selection shapes sequence diversity across the genome of an organism.

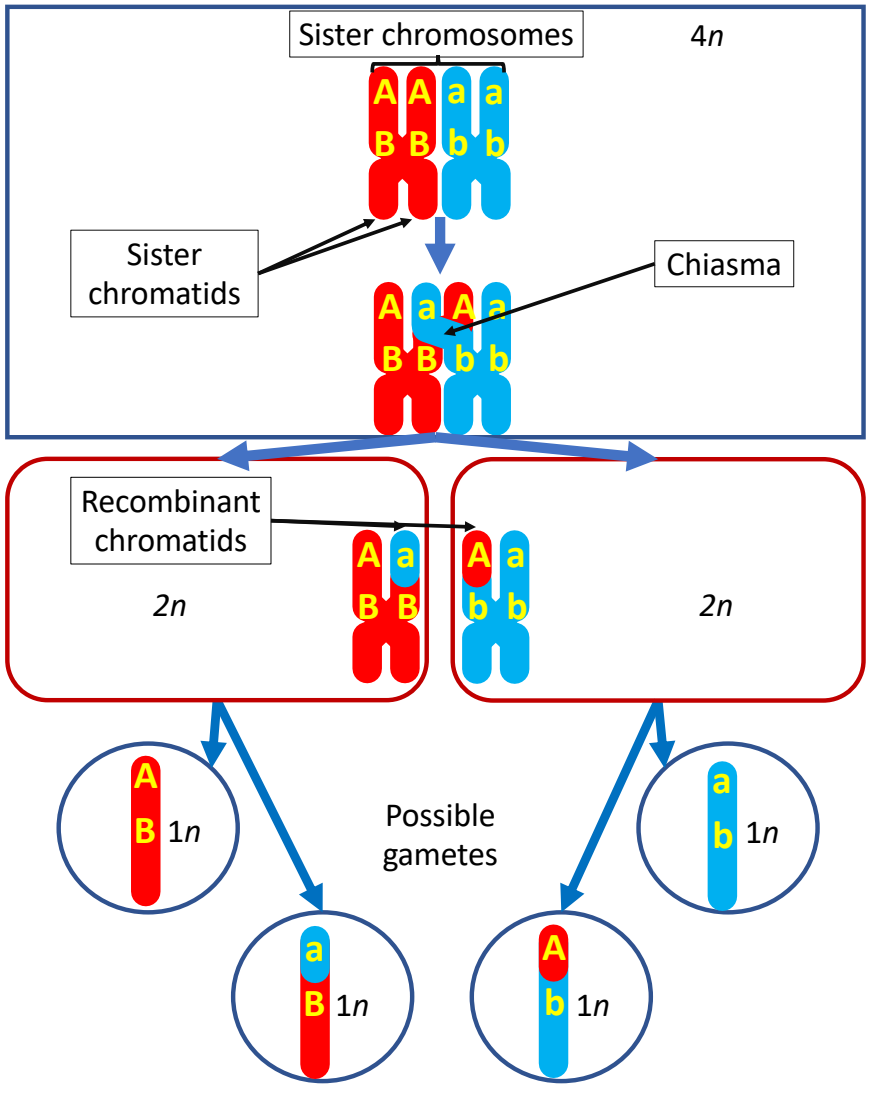

Figure 1. Simplified end of meiosis with cross-over and the produced gametes. Before meiosis, the chromosomes in the cell are duplicated from 2*n* to 4*n*, i.e. each of the two sister chromosomes has sister chromatid. During the first division of meiosis (blue box), the sister chromosomes align together and are attached to each other physically through one or more chiasma. Once the chiasma is resolved, the sister chromosomes are divided to separate cells (the two red boxes). There is a cross-over between the two loci containing A/a and B/b -alleles on one of the chromatids. Finally, when the second division of meiosis completes, single chromatins are separated to gametes containing 1*n* of chromosomes (the four blue circles), which can then combine with gametes from another individual producing a zygote.

1.1.2 Genetic linkage and recombination lead to linkage disequilibrium at the population level

At population level, the consequence of genetic linkage and recombination in chromosomes manifests in linkage disequilibrium (LD). Simply put, LD is the proportion by which two alleles from two different loci along the chromosome appear together more often than expected under completely random mating and random assortment during meiosis (Slatkin 2008). LD constitutes a useful statistic to estimate the independence of pairs of alleles within a study population (Slatkin 2008). LD, both locally and across the whole genome, is usually low in large populations where effective population size is high and where recombination events are likely to lead to novel combinations of alleles. Much of the within chromosome LD is dependent on the same genomic features affecting recombination rates (Smith et al. 2005; Slatkin 2008).

Due to the straightforward link between LD and recombination, patterns of LD can be used to estimate recombination rates (McVean 2004; Chan et al. 2012). LD-based methods for estimating recombination rates have become very popular due to the relative ease by which sequence information can be obtained even from wild populations (McVean 2004; Kulathinal et al. 2008; Silva-Junior & Grattapaglia 2015; Wang et al. 2016; Booker et al. 2017). Traditionally (Morgan 1915), recombination rates have been estimated from relationships of marker positions in linkage maps built based on the marker statuses in offspring resulting from crosses between parents of known genotypes. As hundreds of offspring are required to construct an accurate linkage map, these methods have often been considered cumbersome and are ill suited for application in many species (Nachman 2002; Stapley et al. 2017), contributing to the increasing popularity of LD-based (LDB) methods for estimating recombination rates. Using LD data to infer recombination rates is not without problems, however, as a set of markers is used to infer recombination rates and the inferred recombination rates are then often used to explain patterns of diversity in the same population and often within the same set of markers. Simulations and studies performed using animal model species such as house mouse (*Mus musculus*) (Booker et al. 2017) and in a number of fruit fly (*Drosophila*) species (Kulathinal et al. 2008; Chan et al. 2012) suggest that LDB recombination estimation methods are not strongly affected by natural selection. Less information is available for plant species. As such, further research in these

methods is useful for defining the accuracy of LDB methods compared to more traditional genetic linkage map based (LMB) estimates of recombination in plants (Stapley et al. 2017) (Paper I).

1.1.3 Causes and consequences of linkage disequilibrium

As previously established, levels of LD are usually population specific, and levels of LD are heavily affected by the demographic history of the population as well as historic selective forces that have been applied to the population, both in domesticated (Blair et al. 2018; Serba et al. 2019) and wild species (Flint-Garcia et al. 2003; Slatkin 2008; Lucek & Willi 2021). The extent of LD tends to increase in populations that have been established by limited number of individuals, or have experienced a recent bottleneck, as well as around genomic features that have been under recent selection (Flint-Garcia et al. 2003; Lucek & Willi 2021). Due to this, in wild species, populations at the leading edge of colonization or on the fringes of the species distribution tend to contain overall higher levels of LD compared to more central populations (Lucek & Willi 2021). While selection tends to increase LD in the surroundings of a selected feature (Slatkin 2008), LD also limits the efficiency of both positive (Barton 1995) and negative selection (Felsenstein 1974). Selection in areas of the genome where LD is high can lead to a situation where a selected positive polymorphism "drags" along deleterious polymorphisms to higher frequencies, or even to fixation, within a population. This increase in frequency of deleterious mutations around selected variants is known as "genetic hitchhiking" (Smith & Haigh 1974; Lang et al. 2013), and it has been shown to repeatedly occur during domestication in many plant and animal species, especially close to "domestication genes" that are important for controlling key traits during the domestication process (Schubert et al. 2014; Ramu et al. 2017; Moyers et al. 2018). In some cases, the effect of a selectively favored polymorphism can be countered by linkage to harmful variants (Assaf et al. 2015), severely hampering positive selection on the positive polymorphism. Finally, if multiple positive variants reside within the same linkage block, i.e. part of genome where recombination rarely happens, these variants will be in direct competition, and over time only one of them is likely to be fixed. This phenomenon is termed "Hill-Robertson interference" and it can severely hamper evolution and adaptation in populations with high LD (Hill & Robertson 1966). LD is thus both affected by and affects selection, making

LD important to understand in breeding and conservation contexts (Flint-Garcia et al. 2003; Blair et al. 2018; Serba et al. 2019) (Paper II, Paper IV).

1.1.4 Linkage disequilibrium determines power of genome wide association studies

Linkage disequilibrium along the genome within a population also determines the resolution of association mapping used for dissecting quantitative traits, such as genome wide association studies (GWAS) (Nordborg & Weigel 2008). GWAS is a tool for identifying genetic polymorphisms that are physically linked to loci having a direct causal effect on the phenotype of trait in question, although a GWAS in itself cannot provide evidence of causality for the associations (Pearson & Manolio 2008; Josephs et al. 2017). GWAS have successfully been used to locate candidate genes for a large number of traits across many different species and biological contexts, such as in human (*Homo sapiens*) genetics (Pearson & Manolio 2008), fish conservation genetics (Cauwelier et al. 2018; Waters et al. 2018) and plant breeding (Josephs et al. 2017; Liu & Yan 2019). GWAS has thus become a staple method across the many fields of genetics.

Following the previously established association between high effective population size and low levels of linkage disequilibrium, the resolution and robustness of a GWAS depends on the structure and size of the population it is applied to as well as on the genetic architecture of the trait under study. In terms of genetic architecture, large effect loci with high enough allele frequency are the easiest to detect, while small effect loci with low allele frequencies are likely to be missed in a GWAS. This difficulty of detecting small effect alleles arises from low statistical power due to the weak effect such an allele has on the phenotype. Furthermore, multiple combinations of alleles, i.e. genetic heterogeneity, can lead to a similar phenotype making it hard to identify a given causal allele for a given phenotype. Also, alleles with small effects are more difficult to separate from random effects caused by the environment and other variables (Josephs et al. 2017). Traits affected by limited number of loci with large effects are therefore easiest to study with GWAS (Josephs et al. 2017; Visscher et al. 2017). As for populations, a general rule of thumb is that the larger and the more diverse the population is, the more powerful and high resolution the GWAS is. Conversely, structure within the study population, also known as population stratification, such as relatedness between individuals, increases LD and thus lowers the

potential resolution of the study. Furthermore, population structure can lead to false positives, where a polymorphism's association with trait is caused by the genetic and phenotypic similarity between related individuals rather than due to a true association with the trait (Pearson & Manolio 2008; Josephs et al. 2017; Visscher et al. 2017; Liu & Yan 2019) (Paper III).

As GWAS is based heavily on linkage disequilibrium between loci in a population of phenotyped individuals, the optimal number of markers that are linked to the causal locus is one. Thus, the number of markers included in the GWAS is often trimmed based on the levels of LD to achieve a situation where the whole genome is covered with the lowest possible number of unlinked loci (Ye et al. 2019). This trimming of loci, however, removes information as loci are usually not in complete LD and attempts have been made to find methods that can utilize the whole extent of loci in the population (Bonhomme et al. 2019). Another problem arising from the expansion of association studies from candidate gene -based approaches to whole-genome GWAS is the number of markers that are needed to cover the genome at an acceptable resolution. Methods for adjusting the resulting p-values for multiple testing, such as Bonferroni correction and false discovery rate correction have often been deemed too strict for high numbers of loci as well as for identifying low-frequency loci (Fadista et al. 2016; Bonhomme et al. 2019) (Paper III). To resolve these issues, methods such as Lindley score (LS) have been applied in an attempt to limit the number of independent tests without the need for marker trimming in GWAS studies (Bonhomme et al. 2019).

1.2 Breeding in forestry

Farming began some 11,500 years ago in the Fertile Crescent, which covers modern day Iraq, Syria, Lebanon, Palestine, Israel, Jordan and Egypt, as well as parts of Turkey and Iran. Along with farming began plant domestication and, as its consequence, more goal-oriented plant breeding. Some rudimentary forms of selective breeding may have been utilized, or been the side product of intentional planting of seasonally growing plants, by nomadic tribes in the area few centuries earlier (Zeder 2011; Larson et al. 2014). Many of the earliest plant species to be domesticated were food crops, such as wheat (*Triticum aestivum*) and legumes (*Fabaceae*) like lentils (*Lens culinaris*), and in other parts of the world rice (Oryza sativa) (Asia) and

squashes (*Cucurbita*) (South America). However, since then plants such as flax (*Linum usitatissimum*) and cotton species (*Gossypium*) have also been domesticated and bred for the production of non-dietary fibers (Larson et al. 2014), while flowers like roses (*Rosa*) have been heavily bred for ornamental purposes (Raymond et al. 2018). Interestingly enough, one of the earliest companions of mankind since the dawn of tool use in the early hominins, woody plants, have never been domesticated and bred solely for the purpose of wood production despite their importance as fuel and construction material.

Modern forestry has up until very recently relied either on the regenerative capabilities of the forest or on replanting with seeds from the immediate area or from areas slightly south of the planting area to replenish our wood supply (Hosius et al. 2006). In recent years there has been growing interest in forest tree breeding to utilize novel breeding methods, such as genomic prediction, made possible by the increasing access to genome sequencing. This transition to novel methods has been especially notable in the fast-growing tropical and sub-tropical tree species from the genus of *Eucalyptus*, where many promising results have been observed (Grattapaglia 2004; Tan et al. 2017; Suontama et al. 2019). Forestry industries in northern areas of the globe have been slower to adopt these novel breeding methods, but similar explorations into breeding have been conducted in species native to colder climates, such as conifers (Li et al. 1999; Bouffier et al. 2009; Beaulieu et al. 2014) and willows (Rönnberg-Wästljung & Gullberg 1999; Bonosi et al. 2013; Bubner et al. 2018). Though conifers hold a dominant place in forestry in the north, species in the willow family (*Salicaceae*) hold great potential for short rotation forestry and biofuel production due to their fast growth, high adaptability and clonal reproduction (Weih 2004; Sannigrahi et al. 2010) (Paper II, Paper III, Paper IV).

1.2.1 Genomic prediction is a useful tool for tree breeding

Genomic prediction, or genomic selection as it is often called due to the most common use of genomic prediction being as a tool for selection, is a relatively new breeding tool first proposed by Meuwissen et al. (2001). Simply put, genomic prediction utilizes all available SNP markers simultaneously, though large marker sets are often pruned by LD, to produce a model for prediction of the phenotype of an individual based on genotype information alone. Unlike GWAS, genomic prediction does not directly

provide information on the location of causal polymorphisms (Meuwissen et al. 2001). Instead, data on genome-wide genotypes are combined with phenotype information to construct a statistical model that relates genotypic and phenotypic variation. This statistical model can then be utilized to predict the phenotype of individuals where only the genotype information is available. Genomic breeding values (GEBV) are estimated for each genotyped and phenotyped individual and can be used to estimate the deviation of the progenies from the average of the reference population (Falconer 1981), which is usually the population of potential parents. Aside from the genotype and phenotype of individuals, many of the genomic selection tools also allow inclusion of information on various covariates (Yin et al. 2020). Covariates can be included to further control for the effects of variables with known effect on the phenotype, such as treatments or varying environments. Genomic prediction is particularly suitable for use in forest trees with long generation times, for productivity traits determined by many loci with small effects and for late-manifesting traits (Tan et al. 2017), as genotyping can be done in young individuals without the need for extensive field testing to identify the desirable individuals. Furthermore, it also allows smaller populations of individuals to be kept as genetically undesirable individuals can be pruned out during the early stages of development, partially solving few of the long-standing challenges that have thus far faced tree breeding.

1.2.2 Late manifesting traits, large sizes and local adaptation cause issues in tree breeding

Compared to many other plants, trees pose certain problems unique to them when it comes to breeding. These problems have, in part, led to the slow adoption of modern breeding methods in the forestry industry. Long generation and breeding cycle time of most trees, as well as the late development of many desired traits, has long been a problem, especially under the northern conditions, where growth rates are lower than in equatorial regions. These problems have, in part, been solved by genomic prediction (Tan et al. 2017), which allows for a considerable shortening of the breeding cycle through the ability to screen for desirable genotypes at young age, as alluded to previously.

Issues most relevant to the work in the papers constituting this thesis arise from the high space requirements trees have, limited genetic variability in

small breeding populations (Moran et al. 1980) and the extensive local adaptation many species with wide distributions exhibit (Savolainen et al. 2007; Zhang et al. 2019; Ingvarsson & Bernhardsson 2020). Trees grow considerably larger than the weedy and bushy plants constituting the majority of species domesticated by humans, but potentially require no fewer individuals for a feasible breeding population. While genomic prediction has alleviated the need for space with the ability to select desirable individuals based on their genome alone as previously indicated (Tan et al. 2017), active breeding still requires considerable spatial resources, contributing to high cost of breeding in forest trees (Paper II, Paper III, Paper IV). Furthermore, genetic issues arise from small population breeding causing concerns over sufficient genetic variability of improved seeds and individuals (Moran et al. 1980) (Paper II, Paper IV). Inbreeding is a serious issue in most outcrossing species (Charlesworth & Willis 2009) (Paper IV) and a lack of genetic diversity and heightened LD may also restrict adaptability (Paper II) through multiple means such as Hill-Robertson interference. Furthermore, widespread tree species, such as poplars and aspens (genus *Populus*) (Zhang et al. 2019; Ingvarsson & Bernhardsson 2020), exhibit extensive local adaptation to environmental conditions across their range, and as result, the choice of individuals to include for adaptive breeding to novel conditions is a key issue (Paper II).

1.3 Local adaptation to climate and photoperiod

Local adaptation is a phenomenon where populations of the same species located in different environmental conditions display signatures of phenotypic and genetic adaptation to these specific conditions. These adaptations usually lead to increased fitness for individuals native to the environment compared to individuals that have immigrated into the environment, thus maintaining local adaptation even in populations with considerable gene flow (Figure 2) (Howe et al. 2003; Kawecki & Ebert 2004; Savolainen et al. 2007). Similarly, many key life history traits are often affected by local adaptation, such as growth, predation success or disease resistance (Kawecki & Ebert 2004; Savolainen et al. 2007; Leimu & Fischer 2008). Local adaptation is a common feature in wild populations of widespread species spanning a multitude of different conditions in their distribution (Leimu & Fischer 2008).

Climate and photoperiod are some of the most limiting conditions when it comes to the natural distribution of species along the south-north axis. Large number of species with wide south-north distribution, including humans (Hallmark et al. 2019), thale cress (*Arabidopsis thaliana*) (Fournier-Level et al. 2011), Norway spruce (*Picea abies*) (Beuker 1994) and European aspen (*Populus tremula*) (Luquez et al. 2008), have been observed to exhibit considerable adaptations related to local climate and photoperiod when populations originating at different latitudes were compared. The genetic basis of local adaptation can be highly polygenic with majority of loci conferring small effects to the phenotype (Rockman 2012). However, large effect loci have also been observed in some species (Wang et al. 2018). The polygenic nature of climate and photoperiod adaptation can result in a genetic architecture of local adaptation to climate being very diverse among related species even if they grow under very similar conditions (McKown et al. 2014b; Wang et al. 2018).

At northern latitudes, local adaptation to climate and photoperiod is primarily driven by the need to avoid unfavorable winter conditions with low temperatures, frosts, and poor light conditions. This avoidance is most often manifested in the timing of entering and exiting winter dormancy (referred to as dormancy hereafter) (Doorenbos 1953). Transition from active growth to dormancy and the resumption of growth following dormancy are in many species controlled by differing environmental cues. Entry to dormancy is often induced by changes in photoperiod (Fracheboud et al. 2009) or light quality (Clapham et al. 1998), while release of dormancy is often induced by vernalisation, i.e. a prolonged exposure to low temperatures, followed by an increase in temperature in the spring (Singh et al. 2018). The physical and chemical changes plants exhibit during these transitions are termed autumn phenology, for when the plant prepares to enter dormancy in the autumn, and spring phenology, for when the plant begins active growth following winter dormancy in the spring. Incorrect reaction to the cues signaling the beginning and end of the growing season is likely to lead to loss of potential growth through extension of dormancy into the growing season or loss of realized growth in form of damage to important vegetative tissues. Dormancy represents an important life history trade-off between growth and survival (Loehle 1998). Adequate response to the stimuli marking the beginning and ending of winter is especially important for forest trees, which live under the same conditions for decades or even centuries. Maladapted individuals are

likely to suffer lowered reproductive success and/or biomass production, both of which may have large ecological and economic repercussions (Hurme et al. 1997; Wang et al. 2016) (Paper II, Paper III).

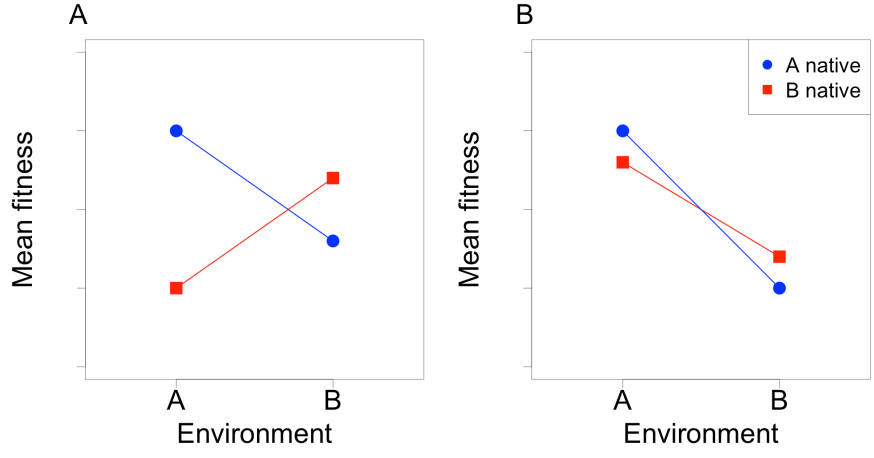

Figure 2. Theoretical examples of local adaptation. A) Both populations have the highest absolute and relative fitness in their native environments. B) Both populations have the highest absolute fitness in environment A, but each have the highest relative fitness at their native environments. This situation is more common in northern climate adaptation, where northern populations tend to have higher absolute, but not relative, fitness in southern environments. Figure adapted from Savolainen et al. (2007).

1.4 Effects of deleterious load in populations

As previously established, the majority of functional mutations, i.e. mutations that affect the amino acid sequence of a protein or affect regulatory elements, are harmful to the function of an organism. In 1935 J.B.S. Haldane posited that harmful or deleterious mutations have negative effects on organism fitness, i.e. the ability of an individual to produce viable offspring throughout its lifetime (Haldane 1935, 2004). With this position, the concept of the phenomenon of deleterious or genetic load was born and has since been extensively studied in a wide range of organisms (Schubert et al. 2014; Henn et al. 2015; Marsden et al. 2016; Zhang et al. 2016; Ramu et al. 2017; Willi et al. 2018). A simple way to define the deleterious load is as all segregating and fixed alleles that negatively affect the fitness of an individual or population at a specific point in time. Deleterious load is thus time and environment dependent phenomenon as alleles that lower fitness in one

environment or point in time may not do the same in another. This is especially true for species with broad distribution ranges, such as forest trees like poplar and aspen species, in which local adaptation is prevalent (Wang et al. 2018; Zhang et al. 2019; Ingvarsson & Bernhardsson 2020). Local adaptation to climate and light conditions clearly exemplify the time dependent aspect of deleterious alleles in the current changing climate conditions, as many populations are at risk of rapidly becoming maladapted to the conditions they historically thrived in (Gougherty et al. 2021), i.e. the once beneficial variants are now becoming deleterious.

Natural selection, and more specifically negative selection, is the main force purging deleterious load. As such, high rates of effective recombination, and as consequence low LD, across the genome is a prerequisite for efficient purging of deleterious alleles from populations (Hill & Robertson 1966; Castellano et al. 2016). Deleterious alleles therefore tend to accumulate in small and structured populations, where the background negative selection is hampered by increased LD (Agrawal & Whitlock 2012; Castellano et al. 2016), though the effects of low effective population size have not been detected in some species (Do et al. 2015). Increases in deleterious load can even lead into mutational meltdown, where the affected population becomes increasingly hampered by its increasing deleterious load, ultimately resulting in population extinction (Lynch et al. 1993; Ramu et al. 2017). This has large implications for breeding populations and small populations of endangered animals, as increasing deleterious load can lead to unexpected results in both breeding and conservation genetics. These unexpected results can potentially be avoided with a careful consideration of the deleterious load. However, our understanding of the behavior of deleterious load during the early stages of active breeding is lacking (Paper IV).

Deleterious load has been suggested as the potential cause behind missing heritability (Hemani et al. 2013), hybrid vigor (Crow 1948; Chen 2010) and inbreeding depression (Charlesworth & Willis 2009), the last of which is a direct contributor to the mutational meltdown discussed earlier. Both missing heritability and inbreeding depression have been suggested to arise from accumulation and fixation of deleterious alleles leading to higher levels of deleterious homozygosity. This happens most often in populations with insufficient genetic variation or which consist of highly related individuals (Charlesworth & Willis 2009; Hemani et al. 2013). Hybrid vigor has been

suggested to, in part, arise from masking of recessive deleterious heterozygote loci in offspring of sufficiently genetically differentiated individuals (Crow 1948; Chen 2010). Inbreeding depression has been posited to be especially dangerous for highly outcrossing species such as maize (*Zea mays*) (Shull 1908; Chen 2010) and is therefore likely dangerous for forest tree species, such as poplars and aspens (Bradshaw & Stettler 1994, 1995). Hybrid vigor has been thoroughly explored both in maize, with the top performing cultivars often being inbred line hybrids (Duvick 2001; Chen 2010) and in hybrid aspen, the interspecies cross between European aspen and American aspen (*Populus tremuloides*), which is one of the most well-established cases of hybrid vigor (Einspahr 1984). It is very important to explore the effects of deleterious load on the viability in these highly outcrossing species, as avoidance of fixation of deleterious alleles might allow for similar results without the need for hybridization (Paper IV).

As implied by the role deleterious load plays in phenomena such as inbreeding depression and hybrid vigor, and the previously established sensitivity to time and population, deleterious load is heavily affected by demographic history and selection through linkage to selected sites. These effects are especially evident in domesticated species, which have gone through strong selection and repeated bottlenecks during their domestication. Studies in species such as dogs (*Canis familiaris*) (Marsden et al. 2016), horses (*Equus ferus*) (Schubert et al. 2014) and cassava (*Manihot esculenta*) (Ramu et al. 2017) all show clear increases in deleterious load compared to their wild relatives. Furthermore, these species, like many others, have specific domestication genes, where the deleterious load is further increased due to the linkage of deleterious alleles to the alleles controlling the desirable domestication trait that has been under intensive selection. Finally, in wild populations, the deleterious load generally increases with distance to the species or population origin as consequence of serial bottlenecks and founder effect, as has been shown in species like humans (Henn et al. 2016) and thale cress (Willi et al. 2018).

Owing to the difficulty in measuring the fitness of an individual, let alone the effect of a variant on it, most methods for defining deleterious loci and alleles rely on more practical forms of defining putative deleterious variants (Kumar et al. 2009; Choi et al. 2015). Methods such as multiple species alignments of exome sequences (genes) (Kono et al. 2018) have been developed. While they are hardly sensitive to the full intricacies of

deleterious load, they serve as a valuable starting point for further research into deleterious load (Paper IV).

1.4.1 Epistatic effects may modulate deleteriousness of damaging alleles

Genes and their products are organized into interacting, interconnected networks within even the simplest species with considerable differences in levels of interconnectivity between different genes (Figure 3) (Brazhnik et al. 2002). This interconnectivity is also in the base of epistasis, where two or more physically unlinked alleles have additional composite effect when present in an individual in certain combinations. The organization of genes to gene networks, epistasis and differential connectivity of genes in gene networks would suggest that not all deleterious mutations or combinations of deleterious alleles have the same impact on a trait or fitness as a whole. It would be prudent to assume that two deleterious mutations in separate gene network or epistatic interaction might not be as harmful as two deleterious mutations in the same gene network (Figure 3). Furthermore, deleterious mutations in highly interconnected genes within the network would likely be more damaging than mutations in non-interconnected genes (Figure 3) (Paper IV). Some of the effects of hybrid vigor (Goodnight 1999; Li et al. 2001), inbreeding depression (Lynch 1991; Li et al. 2001) and missing heritability (Hemani et al. 2013) have been attributed to the gain and loss of epistatic interactions. If loss of positive epistatic effects can affect a trait, so should gaining of negative epistatic effects through deleterious alleles. Epistasis has been often studied in terms of positive epistasis as seen in domesticated pigs (*Sus scrofa*) (Banerjee et al. 2020) and rice (Li et al. 2001), and while studies into epistatic effects of deleterious alleles have been performed, they are less common (Corbett-Detig et al. 2013; Chae et al. 2014). Studies into the epistatic effects of deleterious load are likely to shed light both on the nature of deleterious load as well as on phenomena like inbreeding depression (Paper IV).

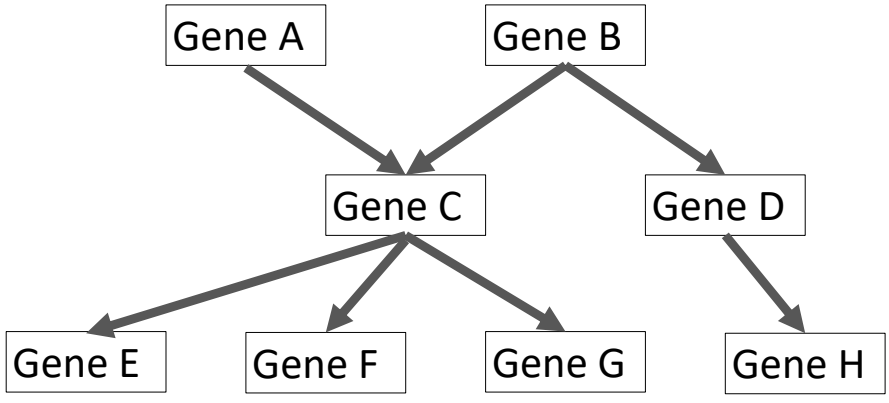

Figure 3. Theoretical gene network. Harmful mutation in gene B is likely more damaging than mutation in gene C, as there are more genes downstream of C (following the arrows) that would be affected. Harmful mutations in genes A and C would likely be more damaging than mutations in genes A and D due to them being in the same pathway (following the arrows).

1.5 Poplar and aspen species

Poplar and aspen species are large, deciduous, early succession trees with wide distributions across the northern hemisphere, spanning from the equator to the northern limits of tree growth. Due to their fast growth, wide distribution ranges as well as the manageable genome size of ~480 Mbp (distributed across a haploid set of 19 ($2n$=38) chromosomes, poplar and aspen species have been established as the *de facto* model system for woody perennials (Taylor 2002; Lin et al. 2018). Both large and small scale synteny is highly conserved among the species in the genus, making many genetic resources easily comparable and transferable between species within the genus (Jansson & Douglas 2007). This has led to many of the species in the genus being fully sequenced, including desert poplar (*Populus euphratica*) (Ma et al. 2013), European aspen (Lin et al. 2018) and the first ever tree species to be sequenced, black cottonwood (Tuskan et al. 2006). Other genomic resources, such as linkage maps, have also been produced for many of the species in the genus (Paolucci et al. 2010; Tong et al. 2016).

Along with scientific interest, many species in the genus are keystone species in their natural habitats (Kouki et al. 2004). As the poplar and aspen species are generally fast-growing, they have great ecological significance as pioneer species, establishing quickly in clearings (Cronk 2005).

Furthermore, the extensive root networks of poplar and aspen species are long lived and extremely resistant to forest fires (Greene & Johnson 1999). Thus, poplar species are among the first species to emerge after disruptions in forest canopy coverage (Greene & Johnson 1999; Cronk 2005). The soft wood of both live and dead trunks and the high amounts of yearly leaf detritus characteristic of poplar and aspen species provide habitats for numerous species of invertebrates, fungi, microbes and vertebrates, including many endangered species of woodpeckers, beetles and fungi. The bark, wood and young shoots of poplar and aspen species provide nourishment to many rodents, such as beavers (*Castor*) and hares (*Lepus*), as well as large ruminants such as moose (*Alces alces*) and white-tailed deer (*Odocoileus virginianus*) (Kouki et al. 2004).

The rapid growth rate and ability to produce natural clones seen in poplar and aspen species has also spurred interest in the forest industry (Taylor 2002; Lin et al. 2018). Species of the genus are frequently utilized in bioenergy production (Sannigrahi et al. 2010) and short rotation forestry (Weih 2004), even outside of their natural distribution ranges (Dickmann & Kuzovkina, 2014). In northern European forestry, poplar and aspen species are an underutilized option due to the phenological maladaptation of commercially bred varieties (Karacic et al. 2003). Commercial interest thus exists for adapting non-native poplar and aspen species to growth under northern European conditions.

1.5.1 European aspen

European aspen is a species of special interest among poplars and aspens as it has the largest distribution of any tree species in Eurasia, spanning from Spain and Scotland in the west to pacific China and Russia in the east, Iceland and northern Scandinavia in the north to northern Africa and southern China in the south (Luquez et al. 2008). The species is present in early succession forests in most forest types present in the boreal and temperate regions and serves as a keystone species in many of these forest types. The extensive distribution of European aspen has been facilitated by strong local adaptation in the populations of the species, despite high rates of outcrossing and considerable gene flow between populations (Wang et al. 2018). This makes the species very interesting in terms of studying local adaptation and spatially varying selection (Luquez et al. 2008; Wang et al. 2018). Finally, the highly outcrossing nature of the species has led to

extensive effective population sizes and consequently high rates of recombination, making the species interesting to study in terms of recombination rate variation across the genome (Paper I).

1.5.2 Black cottonwood

One of the economically interesting species in the genus containing poplars and apsens is black cottonwood, a deciduous tree native to North America with a natural range spanning from Mexico to Alaska in central and northwest North America. The species displays clear signatures of local adaptation to climate and photoperiod in its natural range (Evans et al. 2014; McKown et al. 2014b), and substantial genetic variation across several key adaptive traits with considerable heritability in these traits (Bradshaw & Stettler 1995; Frewen et al. 2000; Yu et al. 2001). Black cottonwood is underutilized in northern European forestry due to phenological maladaptation of commercially bred varieties (Karacic et al. 2003). This has led to breeding programs aiming for climate adapted poplars for northern European climate and light conditions, but these programs suffer from the previously established problems in tree breeding such as space constraints, limited diversity and deleterious load problems (Paper II, Paper III, Paper IV).

2. Aims of the study

I Explore the relationship between LD-based estimates of recombination and genetic linkage map-based estimates of recombination to conclude whether LD-based estimates are sufficient for estimation of recombination rates. Also to assess the effect of different genomic features on recombination rate variation.

II Quantify the genetic diversity in key traits for climatic and photoperiodic adaptation in a small and structured population of black cottonwood, and determine whether the population contains enough genetic variability for adaptation to novel Swedish climate and light conditions.

III Identify candidate genes for an extended set of key traits for climatic and photoperiodic adaptation selection in the small and structured population of black cottonwood.

IV Quantify the deleterious load contained in the small and structured population of black cottonwood and quantify the effects of deleterious load on growth in the population.

3. Results and Discussion

3.1 Linkage disequilibrium-based estimates for recombination rates align accurately with genetic linkage map-based estimates

As established in the introduction of this thesis, meiotic recombination has a considerable role in determining LD, which in turn determines the applicability of tools like GWAS and the effects of selection making it important to understand the recombination landscape of a species. Genetic linkage map based (LMB) estimates of recombination were not available for our black cottonwood study population, forcing us to utilize the LD-based (LDB) estimation of recombination rates to assess recombination rate variation across the genome. However, no information on how LDB and LMB estimates are associated within poplar and aspen species was available. At the same time, LDB estimates had already been used in previous work in the genus within our research group, and as such, the work would also serve to confirm or dispute the validity of these estimates. Consequently, it was important for us to study the association between LDB and LMB estimates among the poplar and aspen species and determine whether LDB methods are feasible to use in the genus. For this, we utilized European aspen, where we had recently produced a high-resolution genetic linkage map, thereby allowing us to make comparisons between different recombination rate estimation methods.

LDB estimates were acquired using LDhelmet (Chan et al. 2012), which is an update of LDhat (Auton & McVean 2007), which had previously been utilized for estimating recombination rates in previous work in the group (Wang et al. 2016). Both LMB and the LDhelmet LDB recombination

estimates showed substantial variation across the genome on a 1 Mbp scale within European aspen at a similar scale with previously observed rates in species such as thale cress (Giraut et al. 2011), black cottonwood (Slavov et al. 2012) and the eucalyptus species rose gum (*Eucalyptus grandis*) (Silva-Junior & Grattapaglia 2015). We also observed a small number of windows where the LDB recombination rates were between 1.5 and 15 times higher or lower than the corresponding LMB rates, which could indicate the presence of recombination hot and coldspots in these windows. Recombination hot and coldspots have been observed in a wide variety of species including thale cress (Kim et al. 2007), maize (He & Dooner 2009) and rice (Si et al. 2015). As recombination hot and coldspots usually span only few kbp (Choi & Henderson 2015), it could be the reason why the relatively coarse LMB estimates fail to detect them whereas they are detectable using the high resolution LDB estimates.

Overall, recombination rate estimates from the consensus linkage map and from polymorphism data show a moderately strong positive correlation of 0.48, which is in a similar scale to correlations between LMB and LDB recombination estimates observed in house mouse (Booker et al. 2017). This suggests that LDB recombination estimates are reliable substitutes for LMB estimates among poplar and aspen species, thus validating the previous work. However, the considerably higher correlations of ~0.8 estimated in threespine stickleback fish (*Gasterosteus aculeatus*) (Shanfelter et al. 2019) indicate that species are highly variable in these correlations, highlighting the importance of species or genus level estimates. Further support for our conclusion that LDB estimates are sufficient proxies for LMB estimates arise from the very similar correlations we observe for these two different estimates with multiple genomic features (Table 1). We observed strong positive correlations between the two recombination rate estimates and gene density (Table 1), which is in line with earlier observations in plants (Tiley & Burleigh 2015; Stapley et al. 2017). This implies that recombination is linked to gene-dense regions, which is likely due to more efficient recruitment of the recombination machinery to euchromatic genome regions, which has also been put forward as an explanation for why recombination rates across plants generally correlate more closely with gene density than physical genome size (Henderson 2012; Tiley & Burleigh 2015). GC-content was also positively correlated with both recombination rate estimates, similar to what has been observed in humans (Fullerton et al. 2001) and thale cress

(Kim et al. 2007) among others. However, this positive correlation is likely indirectly caused by gene density rather than GC-content itself as gene density and GC-content are also highly correlated. Indeed, when GC-content was included in a multiple regression model with gene density, the effect of GC-content was negative for the LMB recombination rate. This is likely due to the higher energy required for strand separation during the strand invasion stage of meiotic recombination in areas with high GC-content due to the higher annealing energy for G-C-nucleotide pairing (Mandel & Marmur 1968). This would explain why GC-content has a direct negative effect on recombination rates when effects of gene density are accounted for. Finally, there were negative correlations between local recombination rates and both repeat density and methylation (Table 1), which are also highly correlated with one another as methylation is more common in regions with repeats. These results are in line with earlier findings of chromatin features' role in establishing crossover locations in plants (Choi et al. 2013; Marand et al. 2019). It has been shown that methylation is lower at observed sites of crossovers (Choi et al. 2013) and cross-over density in maize is negatively correlated with repeats and methylation (Rodgers-Melnick et al. 2015).

Table 1. Summary of LMB and LDB correlations with key genomic features. Highly correlated (>0.80) features not included.

	Gene density	**GC-content**	**Neutral diversity**	**New substitution density**	**Methylation level**
LMB	0.41	0.39	0.45	-0.06	-0.51
LDB	0.35	0.33	0.44	-0.07	-0.35

Both of our recombination rate estimates were strongly correlated with neutral nucleotide diversity (Table 1). It is likely that there exists an interaction between recombination and natural selection, where more neutral diversity is retained in regions with high recombination rates since the effects of linked selection will be reduced in such regions (Begun & Aquadro 1992). A positive correlation between recombination rate and genetic diversity could alternately arise from recombination itself if it was mutagenic (Begun & Aquadro 1992) but in that case we would also expect a positive correlation between recombination rate and sequence divergence at neutral sites (Begun & Aquadro 1992). We did not observe such a correlation in our data as the

correlation observed was weak and negative (Table 1). Linked selection thus has pervasive effects on neutral diversity across the European aspen genome (Ingvarsson 2010; Wang et al. 2016), and this is likely true for the majority of species in the poplar and aspen genus.

3.1.1 The genetic linkage map -based estimates of recombination

To be able to compare the LDB estimates and LMB estimates, we produced marker-dense genetic maps for European aspen utilizing a cross between two northern Swedish individuals. Both the male and female maps were similar in size, 4,054 cM and 4,073 cM respectively. The genetic maps in other poplar and aspen species in the genus have been estimated to be 1,600-3,500 cM long in previous studies (Zhang & Li 2004; Paolucci et al. 2010; Zhigunov et al. 2017), with the true genetic map length estimated to be around 1,800-2,000 cM in the genus. Most studies, including ours, are overestimating the genetic map sizes of poplar and aspen species quite considerably. This is likely caused by the high numbers of markers, which increase the risk of genotyping errors that results in inflation of map sizes as errors are interpreted as recombination events. For comparison, the most recent and high-density genetic map for European aspen with total map length of 3,000-3,100 cM was produced from only 2,000 markers observed in 122 progenies. Our maps were much denser with 12,000-13,000 informative markers observed in 764 offspring, making our framework maps the densest and highest resolution genetic maps available for European aspen at the time, despite the obvious size overestimation. Our framework genetic maps are similar in size to the ~4,200 cM and ~3,800 cM maps for more distantly related species of Eastern cottonwood (*Populus deltoides*) and Chinese cottonwood (*Populus simonii*) respectively (Tong et al. 2016). For these species, it was suggested that the large size was a consequence of inflation resulting from the difficulty of ordering a large number of markers within a linkage group (Tong et al. 2016). Interestingly, while our maps were likely inflated, this had not happened to the same extent as the Eastern cottonwood and Chinese cottonwood genetic maps (Tong et al. 2016) despite considerably larger number of markers used to build them. One explanation for this could be the greater number of segregating progenies utilized for creation of our maps, which likely lowered the impact of random genotyping errors on the map length. Despite the obvious issues of the map sizes, our

maps were robust and worked perfectly for the comparison between LDB and LMB estimates.

3.2 Enough genetic diversity exists in key adaptive traits to accommodate adaptation to novel conditions in small and structured population of black cottonwood

When formulating a breeding plan, some of the first questions potential breeders should ask themselves is whether the traits they are selecting for are sufficiently genetically determined, or heritable, and independent from one another genetically. Furthermore, a breeder needs to confirm their breeding population has enough genetic diversity for these selected traits to enable reaching of the goal of the breeding program. Since we are working with a small and structured population of forest trees, where both the size and the structure pose issues for upkeeping sufficiently sized and sufficiently genetically robust breeding population, the latter question is of special importance for us. Furthermore, understanding the genetic component of the phenotype in each of the selected traits and exploring any potential genetic correlations existing between traits in a multi-trait breeding program is important to avoid issues caused by lack of heritability in the traits or by undetected genetic correlations. The main aim for this breeding program is to establish a climate adapted population of black cottonwood under the novel Swedish conditions. While many of the traits presented here have been explored in the native range of the species, the size of genetic component, genetic variability and the extent of genetic correlations are all dependent on both the population and environment in which they are measured. The traits explored in this part of the study were autumn and spring phenology, i.e. the transition trees make from growth to dormancy in the autumn and vice versa in the spring and the physical changes that occur with these transitions. More specifically leaf color change (yellowing) and leaf drop, or leaf senescence as they are often called constituted proxy traits for autumn phenology transition and bud burst constituted proxy for spring phenology transition. Aside from phenology, we observed lifetime growth and length of growing season, which was defined as canopy duration. The phenology traits were measured in two years (2017-2018), while growth was measured only once (2017)

Between 50-80% of the observed variability in these phenology and growth traits were putatively under additive genetic control as inferred from the narrow sense heritabilities of these traits. This means that the differences observed between individuals are in large part genetic in nature and not caused by random effects such as variations in the microenvironment of each individual in the field. Artificial selection is therefore likely to yield results. Furthermore, considerable genetic variation exists for the phenology traits. These findings are in line with previous studies, where moderate heritabilities have been observed for spring and autumn phenology traits in black cottonwood and other poplar and aspen species. It is worth noting, however, that most of these studies have been conducted in juvenile trees (McKown et al. 2014a; Pliura et al. 2014). This is especially relevant for growth traits, as heritability for growth presented in this study was higher than previously reported for the species (Marron et al. 2010; McKown et al. 2014a), which likely results from the estimation being performed in established adult trees in our study opposed to juveniles as in many other studies. Performing estimation of lifetime growth in adults likely reduces the impact of random events on the heritability leading to higher heritabilites commonly to be found in older individuals (Christensen et al. 2003; Wilson et al. 2005). Other effects, such as the genetic structure in this population may also affect the heritability values. As a result, these findings should be interpreted in the context of the study population.

There were strong genetic correlations of 0.77-0.87 among different spring and autumn phenology traits, which would suggest a shared genetic control underlying both leaf development in the spring and leaf senescence in the autumn and should be taken into consideration when applying selective pressures to these traits. No correlations were found between autumn and spring phenology traits, providing further evidence that the genes and environmental cues underlying spring and autumn phenology are different (McKown et al. 2014b; Singh et al. 2018). This indicates that both spring and autumn phenology can be adjusted independently of one another when utilizing selection. Both initiation and completion of bud burst had negative correlations of -0.32 and -0.42 respectively with lifetime growth suggesting that earlier budburst is beneficial for lifetime growth. As there was a stronger correlation between completion of bud burst and growth, it would suggest that late leaf unfurling, which happens at late stages of bud burst, was especially damaging to lifetime growth. These findings would suggest that

there exists a considerable maladaptation to the novel Swedish conditions present in the population and that earlier bud burst is possible to reach through breeding to maximize growth in Sweden.

3.2.1 Phenology affects length of growing season

As established in the introduction of this thesis, well timed entry to and exit from dormancy is important for optimal growth. This is especially true for trees as any advantages or disadvantages conferred by these timings are likely to accumulate across the years. Length of the growing season in deciduous trees, including black cottonwood, is defined by yearly leaf phenology. Active growth can only begin with actively photosynthesizing fully unfurled leaves and ends with leaf senescence, shortly after which active growth ceases. We found that canopy duration, i.e. the time leaves are actively photosynthesising, was negatively correlated with the two spring phenology (-0.67, -0.43) while positively correlated with the three autumn phenology traits (0.64, 0.60, 0.44). This means early bud burst and late leaf drop lead to longer growing season and greater lifetime growth. Similar patterns have been found in previous studies in the species (McKown et al. 2014a). These findings agree with the trade-off between growth and damage avoidance in climate adaptation in deciduous plants. Early bud burst increases risk of exposure to harmful conditions, but may also confer a competitive advantage through extended early season growth and overall longer growing seasons, while late bud burst ensures the survival of important vegetative tissues from spring frosts but risks the individual falling behind in terms of early season growth and overall growth across the growing season. The late bud bursting individuals may also be affected by spring shading, as the individuals bursting earlier can reduce their growth through shading them with their leaf coverage (Yu et al. 2001).

For autumn phenology, there is an increasing risk of cold damage the later bud set is which can have devastating effects for growth and survival, especially in juvenile trees (Howe et al. 2003). A considerable portion of temperate perennial plants utilize photoperiod rather than temperature to predict the coming of unfavourable conditions and this is true for black cottonwood as well. As such, it is likely that trees are preparing for winter long before changes in leaf colour occur (Basler & Körner 2012). Plenty of evidence exists suggesting that temperature is not a trigger for senescence (Keskitalo et al. 2005; Luquez et al. 2008; Fracheboud et al. 2009), but

evidence from hybrid poplars suggests that temperature may contribute to the timing of growth cessation and bud set (Rohde et al. 2011). It is clear, however, that the rate at which senescence progresses is temperature dependent (Fracheboud et al. 2009). Yearly variation in autumn phenology observed between years 2017 and 2018 is therefore likely a response to different temperatures and other conditions, like water availability, suggesting that large between year shifts in autumn phenology timing are possible despite the photoperiod-linked nature of autumn phenology initiation. This suggests that leaf colour and senescence traits may also be responding to multiple environmental cues either during the summer growing season, or during the onset of autumn. We will discuss the role of temperature in phenology more thoroughly in the next chapter.

3.3 Genome wide association study reveals potential selection targets for adaptive breeding in small and structured population of black cottonwood

While the importance of GWAS as a breeding tool has somewhat diminished since the conception of genomic prediction, GWAS still offers important insights into the genetic architecture of traits and indicates where the selective pressures during active breeding may be the heaviest. The traits included in this study, phenology and growth, have been often studied in their native environment and many GWAS have also been performed in the native range (Evans et al. 2014; McKown et al. 2014b), it is important to study the genetic architecture of these traits in the novel environment of Sweden. The change from the native environment of the species to the novel environment in Sweden may have affected the relative levels of importance of the genes part of the genetic architecture of the phenology and growth traits explored in this study as novel stressors may lead into novel alleles being more important as limiting factors for the traits. As previously mentioned, GWAS will shed light on what parts of the genome will be subjected to selective forces, which is important for future planning, for example, to avoid unnecessary increases in the deleterious load within and around the selected genomic areas. GWAS allows us to expand our understanding of the genetic basis of adaptation to the novel conditions of Sweden and to avoid unexpected problems during the active breeding. We included all the phenology and growth traits from the previous chapter in this

study, but also added multiple bud break and bud set traits derived from a phytotron study which gave us access to trait observations from a controlled climate chamber in an effort to reduce effects from environmental variation.

Most of the growth and phenology traits used in this study displayed appreciable levels of heritability, between 19% and 88% for narrow-sense heritability. Only the initiation stages of both bud burst in the top bud and initiation of bud set were lacking of both phenotypic variation and narrow-sense heritability. Heritability of a trait is known to vary between different environments and life stages in the same plant genotype (Christensen et al. 2003; Wilson et al. 2005), though measurements taken in controlled environment tend to be higher (Ali & Johnson 2000), which is opposite to our findings. As bud burst initiation of the top bud and initiation of bud set are often the traits that best capture the timing of exiting and entering dormancy in plants, the lack of phenotypic variation and heritability in these traits is concerning. However, this lack of variability and heritability in these traits is, at least partially, explained by the damage to the apical buds sustained by many individuals during equipment break-down in the phytotron. However, we cannot rule out that it results from previous selection, either artificial or imposed by the novel conditions of Sweden, in the population.

We performed GWAS using over 7 million SNP markers, which is in drastic violation of the 1 marker for 1 segment of genome rule of thumb for GWAS presented before. As such, we utilized a Lindley score -based approach that allowed us to take advantage of the large number of markers available without biasing the results by lowering the statistical power due to the large number of non-independent markers. It also allowed us to better utilize the information contained in the patterns of linkage disequilibrium among adjacent markers and further mitigate the effects of having a small and structured study-population (Bonhomme et al. 2019). LD decays on average over 10 kbp in our study population. As such, we augmented all of the significant genome segments located from the Lindley scores by 10 kbp in either direction (these 10 kbp expanded significant LS segments are referred to as segments henceforth) to identify potential candidate genes. There were 250 significant segments with 584 genes within 10 kbp of these segments across all of the study traits that displayed sufficient phenotypic and genetic variation (Figure 4). Of these 584 candidate genes, 5 were shared between years in spring phenology traits and 11 in autumn phenology traits

(Figure 4). It is these shared genes that provide us the most reliable candidate genes for selection in autumn and spring phenology.

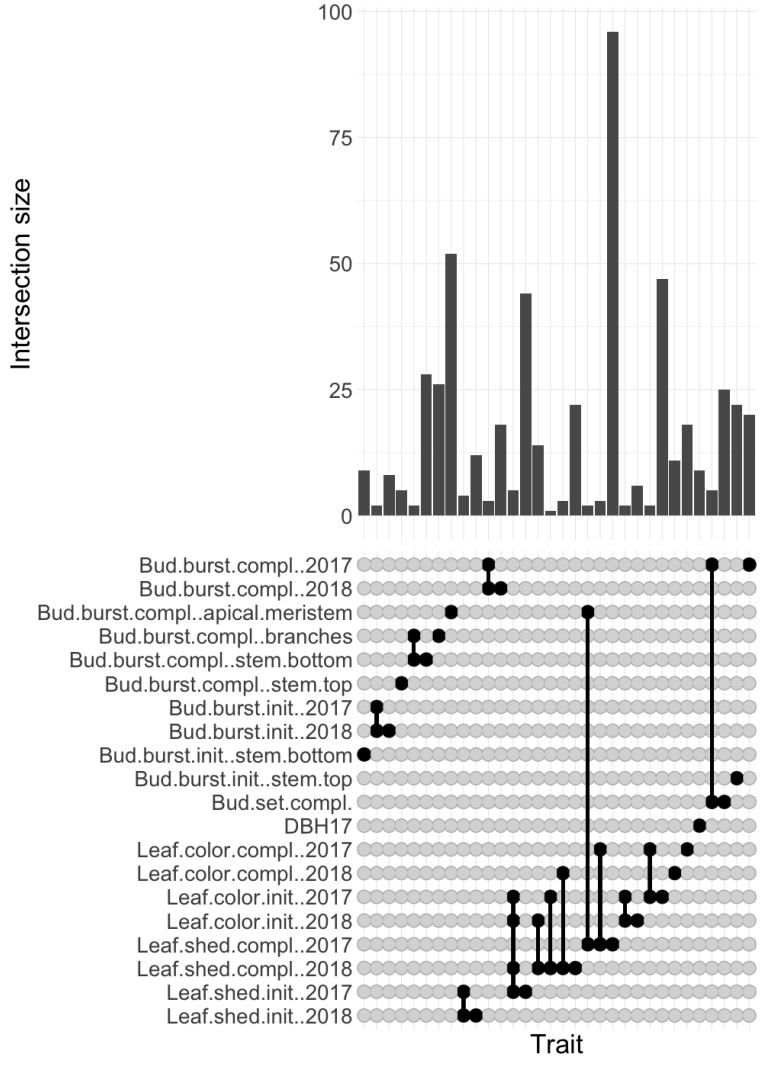

Figure 4. Summary of gene numbers and overlaps for all of the paper III study traits in alphabetical order. When only one dot is blackened in the trait matrix, then the intersection size in the above bar plot equals to the number of genes found in the trait. If multiple dots are blackened and connected with a line in the trait matrix, then the intersection size in the above bar plot equals to the number of genes shared between these traits (Lex et al. 2014).

3.3.1 The two study years might offer different insights for adaptation and selection

As briefly touched upon in the previous chapter, years 2017 and 2018 were considerably different in terms of temperature and precipitation during the study, with 2018 being both considerably warmer and above the 5ºC considered as growing season temperature longer than year 2017. The difference in mean temperature between the two study years is mostly caused by an early and very warm spring and an exceptional heatwave during July in 2018, with other months being less drastically different. 2018 also had lower levels of precipitation across the whole year and the earlier autumn phenology initiation and completion in 2018 may therefore have been brought upon by the lowered water availability as well as the higher temperatures, though water availability was not directly measured.

Previous work has shown temperature to be the main driver of the initiation of spring phenology transitions in many perennial species, including poplar and aspen species (Singh et al. 2018), and higher temperatures have also been found to facilitate quicker completion of spring phenology transitions in other species (Nicotra et al. 2010; Tansey et al. 2017). Furthermore, despite the initiation of autumn phenology being largely determined day length in poplar and aspen species (Fracheboud et al. 2009), evidence exists indicating both temperature and water availability affecting both the timing of initiation and completion of autumn phenology transition, especially when it comes to leaf phenology (Kalcsits et al. 2009; Ghelardini et al. 2014; Xie et al. 2018). Some evidence also exists of earlier spring phenology transition leading to earlier autumn phenology transition (Keenan & Richardson 2015), while water availability and rain patterns have been shown to affect the speed and timing of initiation of autumn phenology, though the direction of these effects seem highly dependent on the species (Xie et al. 2018). Our findings are well in line with these previously established relationships between temperature, rainfall, water availability and phenology timing. Indeed, it is likely that the conditions during year 2018 induced the earlier onset and completion of both spring and autumn phenology transitions in the year, though it still is unclear to us, which of these relationships were the truly causal ones. Furthermore, we identified noticeably fewer significant LS -segments for autumn phenology traits in 2018 than 2017, which could indicate different genetic architecture for autumn phenology in year 2018, though further work is required to prove or

disprove such claims. Nonetheless, the common candidate genes identified for the phenology traits across the two years offer potential targets for adaptive improvement of black cottonwood to Swedish conditions as they are stable across the two very different years, though complementary studies are required to confirm the roles of the specific candidate genes due to the weaknesses of our population.

3.3.2 Spring phenology displays complex genetic architecture

Previous studies have shown that the apical meristem is unique in terms of functions compared to axillary meristems in species with apical dominance, with one of these extra functions being inhibition of axillary buds through apically produced, basipetally transported auxins (Chatfield et al. 2000; Ljung et al. 2001; Qiu et al. 2019). In our study, we timed the bud burst of different buds in our phytotron experiment. The results clearly show the apical dominance effect, as the apical bud is seen to initiate bud burst earlier than the other bud types. At the same time, the buds along the stem have significantly later date of bud burst initiation conditional in presence of undamaged apical bud. As such, it is clear that an undamaged apical bud suppresses the initiation and early stages of bud burst in other buds as has been seen also in other species (Chatfield et al. 2000; Qiu et al. 2019). These findings underline the importance of the apical bud for repressing bud burst of stem buds during spring phenology transitions, while branch buds seem function independently of the apical bud, with no significant difference in timing of bud burst initiation between individuals with undamaged and damaged apical buds.

The higher numbers of significant genes observed in the phytotron combined with the lack of large-scale overlap in candidate genes between the bud burst traits phenotyped in our study (Figure 4), especially within the controlled conditions of the phytotron, seems to not only support the well-established view that bud burst is a highly plastic trait but also that is has a complex genetic basis (Evans et al. 2014; McKown et al. 2014b). The lack of overlap in candidate genes identified between the bud burst traits scored in the different buds in the phytotron experiment further suggests that the genes controlling bud burst in different parts of the plant are considerably different with no large effect genes affecting multiple bud types. This is likely to have an effect on how results both within and between studies should be compared as the buds compared need to be sufficiently similar.

Further work is required to fully understand the extent of differences in genetic control of different bud types. It is also worth keeping in mind that the study population used was small and structured, which required the use of stringent thresholds for Lindley score and as result many weaker associations for bud burst traits could have been filtered out.

Five candidate genes were identified for bud burst in the field, of which 2 and 3 shared between the initiation and completion stages of bud burst respectively (Figure 4). This aligns with the genetic correlations found for these traits previously (Paper II), but would surprisingly suggest a more consistent genetic control in spring phenology between the years in the field than genetic control found in autumn phenology. Comparing with the previously mentioned studies in the native range, we found no overlap between all three studies, but found overlap of 20 genes candidate genes with the less stringent study. This suggests at least a small degree of similarity in the genetic control of bud burst between the native and the Swedish environments, but the overlap is substantially lower than that observed in autumn phenology traits, once again indicating the need for novel allele combinations for adaptation.

3.3.3 Initiation of autumn phenology transitions are more genetically controlled than completion

Initiation of autumn phenology transition has been previously observed to remain largely consistent between conditions and years in the same trees, suggesting stringent genetic control of the initiation of autumn phenology traits (Fracheboud et al. 2009; Rohde et al. 2011) (Paper II). Our findings are in line with this as the 11 candidate genes found shared between the two years in the field for autumn phenology traits were found in the initiation stages of leaf autumn phenology (Figure 4). Autumn phenology traits have also been shown to have a degree of shared genetic control (McKown et al. 2014b) and to be genetically correlated in this population (Paper II). Our results support this as notable overlap in candidate genes between different autumn phenology traits in the field were observed with the largest being 15 candidate gene overlap between initiation of leaf color change and completion of leaf shed (Figure 4). This overlap could be further evidence for a systematic stress response to heat and/or drought experienced in year 2018. Both high temperatures and low water availability have been previously found to speed up autumn phenology in model species (Munné-

Bosch & Alegre 2004; McKown et al. 2014b), which fits well with our observations of the weather conditions in year 2018.

We also identified two candidate genes for autumn phenology that were identified in two earlier studies performed in the native range of the species (Evans et al. 2014; McKown et al. 2014b), both of which have functions in metabolism of complex carbohydrates. They may have roles in many processes important to autumn phenology transitions such as cell wall degradation or production of storage carbohydrates. The existence of these shared genes suggests there to be a number of the larger effect loci for the genetic architecture of autumn phenology traits that are shared across both native and novel environments. However, the relatively large numbers of study specific candidate genes observed in all these studies also hint towards the complexity of these traits under variable natural conditions. There was very little overlap between the two years in our results under the Swedish conditions and the results from the two other studies (Evans et al. 2014; McKown et al. 2014b), clearly displaying the need for unique allele combinations when dealing with novel environmental stresses in order to reach optimal adaptation.

3.4 Deleterious load affects growth and phenotypic gain in a small population of black cottonwood

The potentially catastrophic effects deleterious load can have on a small population were well established in the introduction of this thesis. As our potential breeding population is characterised by LD and contains limited amount of genetic diversity, ignoring deleterious load could risk reduced fitness and ultimately mutational meltdown of the population in the future generations. To avoid this, we decided to explore the possibilities that taking deleterious load into account could offer by exploring the effects of deleterious load on lifetime growth in the population. As previously indicated, lifetime growth is a good proxy for fitness and fitness is far more difficult to measure in black cottonwood than diameter at breast height. We utilized the DBH results also used in papers II and III and shall refer to DBH as growth in this chapter of the thesis.

We utilized multiple methods to uncover the effects that deleterious load has on growth in our population. Among these methods was using the the beta-values for DBH produced in the GWAS performed for paper III. These

values clearly indicated that our putatively deleterious alleles had significantly more negative effects on growth than expected from a random sample of similar size. There was a significant negative correlation between deleterious load and growth in our population, though we observed a stronger correlation between randomly sampled derived homozygotes and lifetime growth. The correlation between random homozygotes and growth was likely caused by the geographic distance from Nisqually-1, the black cottonwood reference individual, which originates from Washington, USA. The southernmost parental individuals in our study population also originated from Washington or neighboring states. As such, the very similar strong correlations between both random (-0.61) and tolerated (-0.65) homozygotes and lifetime growth likely stem from demographic effects. The correlation between lifetime growth and deleterious homozygotes (-0.28) is lower, which could indicate our putative deleterious homozygotes have been selected against across generations. If this is the case, it must be because the deleterious homozygotes damage fitness and are thus selected against. Similar, though less drastic, differences were found for allele counts.

We also observed considerable differences in the amount of deleterious load between individuals, suggesting that controlling the segregating load is possible in the population across the future generations. Following this, genomic prediction utilizing deleterious load information as covariate far outperformed predictions without covariates in both prediction accuracy and variance, implying that important information is contained in the deleterious load of an individual. These improvements could also be due to the extra sequence information contained in the counts, but this seems unlikely given the simulation results discussed next. Inclusion of deleterious allele counts in some form could offer benefits in breeding, especially as defining deleterious alleles becomes more accurate. To further test this, we simulated 20 generations of GEBV-based breeding and observed notable differences in long-term breeding success, or breeding cycle phenotypic gain, depending on whether deleterious load was utilized as covariate in the simulations or not. While both simulations showed a rapid increase in average growth between the parental population and the first offspring population (F1), in generations F2-F20 simulations with deleterious load covariate showed greater gains in growth than simulations with no covariate, where the population mean growth declined to parental levels by F20. These findings are indicative of potential long-term benefits of controlling for the

deleterious load during breeding. While our small population does not allow for comparison with higher number of selected individuals, it seems likely that controlling for deleterious load is especially important in small-scale breeding where deleterious load poses a greater threat (Lynch et al. 1993). Deleterious load itself was very similar at generation F20 in both simulations, which would suggest more complexity within the deleterious load and accounting for this complexity could potentially even further improve both genomic prediction and breeding cycle gain in growth.

3.4.1 Deleterious load associates with multiple genomic features along the genome

We observed considerable variability in deleterious load, i.e. the number of deleterious loci and alleles, in a window-based approach along the genome and this variation largely seems be controlled by the amount of coding sequence in different windows. We saw considerably fewer deleterious loci in the study genes compared to tolerated and amino acid changing loci and the frequency distribution of deleterious alleles was also significantly skewed towards low frequencies compared to both random and tolerated derived alleles, indicating that fewer deleterious alleles reach high frequencies. This is further evidence that negative selection (Hill & Robertson 1966; Castellano et al. 2016) is likely reducing the number and frequencies of the putative deleterious loci compared to their alleles across the genome.

There was a moderate positive correlation (0.32) between deleterious locus number and amount of coding sequence in the window, while a weak but significant (-0.04) negative correlation was observed between deleterious allele frequency and amount of coding sequence. The observed correlations were not especially strong suggesting other features are likely important determinants of deleterious load. Taken together with our findings in the putative centromeric regions reported for black cottonwood (Weighill et al. 2019) we are firm that deleterious alleles are more likely to reach higher frequencies in gene-sparse, centromeric regions as opposed to telomeric regions with higher gene density, while the opposite is true for the number of deleterious loci. It is well known that centromeric regions have lower levels of recombination than telomeric regions (Houben & Schubert 2003; Mancera et al. 2008; Lynch et al. 2010).

In paper I, we showed that recombination rates are higher in gene-dense regions of the genome in the related species European aspen due to more efficient recruitment of recombination machinery in such regions (Henderson 2012; Tiley & Burleigh 2015). Recombination rates were significantly negatively correlated with deleterious locus count (-0.16), deleterious allele count (-0.14, $p<0.01$) and deleterious homozygote count (-0.11), i.e. less deleterious load exists in regions with high recombination, which is in line with previous results in maize (Rodgers-Melnick et al. 2015) and cassava (Ramu et al. 2017). The higher recombination rates in gene-dense areas explain both the higher numbers of deleterious loci retained in gene-dense areas and the low deleterious allele frequencies in these loci through the linkage breaking effects of recombination. Overall, our findings clearly indicate efficient purging of low frequency deleterious variants in high recombination regions.

3.4.2 Epistasis plays a role in determining damage caused by deleterious alleles

As presented in the introduction of this thesis, epistatic interactions between deleterious alleles could explain parts of the phenomena of missing heritability and also inbreeding depression. This could also explain some of the discrepancy observed between deleterious allele and random allele correlations with growth presented earlier. We therefore set out to explore deleterious epistasis in our population. The results of these analyses clearly indicate that deleterious alleles have notable non-additive deleterious effects on growth. The number of significant clusters of epistatically linked deleterious alleles was not very high and many of the clusters contained high numbers of deleterious alleles. These findings support the view presented earlier, that all combinations of deleterious alleles do not have the same direct effect on growth (Elena & Lenski 1997; Sanjuan & Nebot 2008). These epistatic modules could be at the root of the complexity our simulations hinted at.

GO-term analysis of the epistatic clusters showed majority of the clusters played a role in stress, transport or membranes. A few clusters also had roles in growth and cellulose synthesis. These findings suggest that damaging effects of deleterious alleles on growth could often be caused by diminished ability to react to external stressors, such as those seen in field data from 2018, or from a diminished ability to transport materials and signals

efficiently within individuals. The presence of the specific term for root development as well as multiple water availability-based stress gene pathways in multiple modules might further suggest that the root system is especially vulnerable for deleterious effects in the Swedish environment.

4. Conclusions

In the papers constituting this thesis, we have presented genetic linkage map of unprecedented resolution for European aspen. Using this map and a number of genomic features, we showed that recombination rates estimated from LD are feasible alternative for the traditional genetic linkage map - based estimates. We also showed that the small and structured population of black cottonwood used in this study contained sufficient amounts of genetic variation in key traits for adaptation to novel Swedish conditions and that spring and autumn phenologies were genetically uncorrelated. We found a number of candidate genes that could serve as target for selection in these key traits to facilitate adaptation. We also showed that both spring and autumn phenology likely require novel allele combinations for adaptation to Swedish condition. We found distinct, large effect genes controlling bud burst in buds from different parts of the plant and clear indications of a strong apical effect in black cottonwood. Finally, we show that the deleterious load in our population has negative effects on growth. Furthermore, deleterious load can be partitioned into epistatically connected clusters, suggesting that deleterious alleles may have different effects depending on the epistatic environment of the allele. By taking the deleterious load into account during breeding it may be possible to improve genetic gain in traits such as growth. Overall, we show that the small and structured population of black cottonwood described in this thesis could function as the starting point of a commercially viable breeding programme for black cottonwood in Sweden, proving it the little population that could.

5. Future perspectives

Our small black cottonwood population is ready to be used for breeding as we have demonstrated the utility of the population in breeding black cottonwood adapted to Swedish climate. However, during breeding, many of our more theoretical conclusions can be proven or disproven and as such it would be extremely important to study the upcoming generations of this population closely. We would be able to discern whether such a small population truly has the genetic diversity required for adaptation to novel conditions and which genomic regions are subject to the strongest selective effects during breeding. This observation would be especially important for our knowledge of the behavior of the deleterious load. Recording this behavior in beginning of breeding would be invaluable for breeding in similarly small and structured populations elsewhere.

Deleterious load and its epistatic effects are also important to study further. It would be important for application of our findings here to perform similar simulations of deleterious load behavior in larger populations. This would allow us to truly confirm whether consideration of deleterious load is only important to small breeding populations or if the benefits shown in this thesis extend to large-scale breeding as well. Furthermore, we will explore whether partitioning deleterious load by the clustering found in the epistatic network analysis will result in further improvements of predictive ability and breeding cycle gain in the population.

Further work with deleterious load in an extended dataset of European aspen has also been planned. These plans would provide a link to our earlier work in European aspen, where we tested the LDB recombination estimation methods in, and by offering a local, closely related species to perform comparisons with. Furthermore, it will offer a species with a unique colonization history and where we have access to strong, naturally occurring

selective sweeps to compare the effects of sweeps in artificial and natural contexts on deleterious load.

Another avenue of future work of interest would be to expand the link found between deleterious load and positive selection/selective sweeps to negative selection. While we could quite clearly see that the deleterious load had been affected by external forces that were clearly different from the demographic effects seen in sample the random alleles, we could only speculate that this was negative selection. As such, studying what effects negative selection has had on the deleterious load in the population would confirm or disprove our speculations.

References

Agrawal, A.F. & Whitlock, M.C. (2012). Mutation Load: The Fitness of Individuals in Populations Where Deleterious Alleles Are Abundant. *Annual Review of Ecology, Evolution, and Systematics*, 43 (1), 115–135. https://doi.org/10.1146/annurev-ecolsys-110411-160257

Ali, A. & Johnson, D.L. (2000). Heritability estimates for winter hardiness in lentil under natural and controlled conditions. *Plant Breeding*, 119, 283–285

Assaf, Z.J., Petrov, D.A. & Blundell, J.R. (2015). Obstruction of adaptation in diploids by recessive, strongly deleterious alleles. *Proceedings of the National Academy of Sciences*, 112 (20), E2658–E2666. https://doi.org/10.1073/pnas.1424949112

Auton, A. & McVean, G. (2007). Recombination rate estimation in the presence of hotspots. *Genome Research*, 17 (8), 1219–1227. https://doi.org/10.1101/gr.6386707

Banerjee, P., Carmelo, V.A.O. & Kadarmideen, H.N. (2020). Genome-Wide Epistatic Interaction Networks Affecting Feed Efficiency in Duroc and Landrace Pigs. *Frontiers in Genetics*, 11 (February), 1–13. https://doi.org/10.3389/fgene.2020.00121

Barton, N.H. (1995). Linkage and the limits to natural selection. *Genetics*, 140 (2), 821–841. https://doi.org/10.1093/genetics/140.2.821

Basler, D. & Körner, C. (2012). Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agricultural and Forest Meteorology*, 165, 73–81. https://doi.org/10.1016/j.agrformet.2012.06.001

Bateson, William, Saunders, Edith Rebecca, Punnett, Reginald Crundall (1905) Experimental studies in the physiology of heredity. In: Reports to the Evolution Committee of the Royal Society. The Royal Society of London, London.

Beaulieu, J., Doerksen, T.K., MacKay, J., Rainville, A. & Bousquet, J. (2014). Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genomics*, 15 (1), 1–16. https://doi.org/10.1186/1471-2164-15-1048

Begun, D.J. & Aquadro, C.F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*, 356 (6369), 519–520. https://doi.org/10.1038/356519a0

Benson, K.R. (2001). T. H. Morgan's resistance to the chromosome theory. *Nature Reviews Genetics*, 2 (6), 469–474. https://doi.org/10.1038/35076532

Beuker, E. (1994). Adaptation to climatic changes of the timing of bud burst in populations of Pinus sylvestris L. and Picea abies (L.) Karst. *Tree Physiology*, 14 (7-8–9), 961–970. https://doi.org/10.1093/treephys/14.7-8-9.961

Blair, M.W., Cortés, A.J., Farmer, A.D., Huang, W., Ambachew, D., Varma Penmetsa, R., Carrasquilla-Garcia, N., Assefa, T. & Cannon, S.B. (2018). Uneven recombination rate and linkage disequilibrium across a reference SNP map for common bean (Phaseolus vulgaris L.). *PLoS ONE*, 13 (3), 1–21. https://doi.org/10.1371/journal.pone.0189597

Bonhomme, M., Fariello, M.I., Navier, H., Hajri, A., Badis, Y., Miteul, H., Samac, D.A., Dumas, B., Baranger, A., Jacquet, C. & Pilet-Nayel, M.L. (2019). A local score approach improves GWAS resolution and detects minor QTL: application to Medicago truncatula quantitative disease resistance to multiple Aphanomyces euteiches isolates. *Heredity*, 123 (4), 517–531. https://doi.org/10.1038/s41437-019-0235-x

Bonosi, L., Ghelardini, L. & Weih, M. (2013). Towards making willows potential bio-resources in the South: Northern Salix hybrids can cope with warm and dry climate when irrigated. *Biomass and Bioenergy*, 51, 136–144. https://doi.org/10.1016/j.biombioe.2013.01.009

Booker, T.R., Ness, R.W. & Keightley, P.D. (2017). The Recombination Landscape in Wild House Mice Inferred Using Population Genomic Data. *Genetics*, 207 (1), 297–309. https://doi.org/10.1534/genetics.117.300063

Bouffier, L., Raffin, A., Rozenberg, P., Meredieu, C. & Kremer, A. (2009). What are the consequences of growth selection on wood density in the French maritime pine breeding programme? *Tree Genetics and Genomes*, 5 (1), 11–25. https://doi.org/10.1007/s11295-008-0165-x

Bradshaw, H.D. & Stettler, R.F. (1994). Molecular genetics of growth and development in Populus. II. Segregation distortion due to genetic load. *Theoretical and Applied Genetics*, 89 (5), 551–558. https://doi.org/10.1007/BF00222447

Bradshaw, H.D. & Stettler, R.F. (1995). Molecular genetics of growth and development in Populus. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. *Genetics*, 139 (2), 963–973. https://doi.org/10.1093/genetics/139.2.963

Brazhnik, P., De La Fuente, A. & Mendes, P. (2002). Gene networks: How to put the function in genomics. *Trends in Biotechnology*, 20 (11), 467–472. https://doi.org/10.1016/S0167-7799(02)02053-X

Bubner, B., Köhler, A., Zaspel, I., Zander, M., Förster, N., Gloger, J.C., Ulrichs, C. & Schneck, V. (2018). Breeding of multipurpose willows on the basis of Salix daphnoides Vill., Salix purpurea L. And Salix viminalis L. *Landbauforschung*, 68 (1–2), 53–66. https://doi.org/10.3220/LBF1538634874000

Carlin, J. L. (2011) Mutations Are the Raw Materials of Evolution. *Nature Education Knowledge* 3(10):10:Link:https://www.nature.com/scitable/knowledge/library/mutations-are-the-raw-materials-of-evolution-17395346/ Acquired 12.07.2021

Castellano, D., Coronado-Zamora, M., Campos, J.L., Barbadilla, A. & Eyre-Walker, A. (2016). Adaptive evolution is substantially impeded by hill-Robertson interference in drosophila. *Molecular Biology and Evolution*, 33 (2), 442–455. https://doi.org/10.1093/molbev/msv236

Cauwelier, E., Gilbey, J., Sampayo, J., Stradmeyer, L. & Middlemas, S.J. (2018). Identification of a single genomic region associated with seasonal river return timing in adult scottish atlantic salmon (Salmo salar), using a genome-wide association study. *Canadian Journal of Fisheries and Aquatic Sciences*, 75 (9), 1427–1435. https://doi.org/10.1139/cjfas-2017-0293

Chae, E., Bomblies, K., Kim, S.T., Karelina, D., Zaidem, M., Ossowski, S., Martín-Pizarro, C., Laitinen, R.A.E., Rowan, B.A., Tenenboim, H., Lechner, S., Demar, M., Habring-Müller, A., Lanz, C., Rätsch, G. & Weigel, D. (2014). Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. *Cell*, 159 (6), 1341–1351. https://doi.org/10.1016/j.cell.2014.10.049

Chan, A.H., Jenkins, P.A. & Song, Y.S. (2012). Genome-Wide Fine-Scale Recombination Rate Variation in Drosophila melanogaster. *PLoS Genetics*, 8 (12). https://doi.org/10.1371/journal.pgen.1003090

Charlesworth, D. & Willis, J.H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, 10 (11), 783–796. https://doi.org/10.1038/nrg2664

Chatfield, S.P., Stirnberg, P., Forde, B.G. & Leyser, O. (2000). The hormonal regulation of axillary bud growth in Arabidopsis. *Plant Journal*, 24 (2), 159–169. https://doi.org/10.1046/j.1365-313X.2000.00862.x

Chen, J. & Furano, A. V. (2015). Breaking bad: The mutagenic effect of DNA repair. *DNA Repair*, 32 (1), 43–51. https://doi.org/10.1016/j.dnarep.2015.04.012

Chen, J., Miller, B.F. & Furano, A. V (2014). Repair of naturally occurring mismatches can induce mutations in flanking DNA. *eLife*, 3 (3), 1–27. https://doi.org/10.7554/eLife.02001

Chen, Z.J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. *Trends in Plant Science*, 15 (2), 57–71. https://doi.org/10.1016/j.tplants.2009.12.003

Choi, K. & Henderson, I.R. (2015). Meiotic recombination hotspots - A comparative view. *Plant Journal*, 83 (1), 52–61. https://doi.org/10.1111/tpj.12870

Choi, K., Zhao, X., Kelly, K.A., Venn, O., Higgins, J.D., Yelina, N.E., Hardcastle, T.J., Ziolkowski, P.A., Copenhaver, G.P., Franklin, F.C.H., Mcvean, G. & Henderson, I.R. (2013). Arabidopsis meiotic crossover hot spots overlap

with H2A.Z nucleosomes at gene promoters. *Nature Genetics*, 45 (11), 1327–1338. https://doi.org/10.1038/ng.2766

Choi, Y., Chan, A.P. & Craig, T.J. (2015). PROVEAN web server : a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31 (April), 2745–2747. https://doi.org/10.1093/bioinformatics/btv195

Christensen, K., Frederiksen, H., Vaupel, J.W. & McGue, M. (2003). Age trajectories of genetic variance in physical functioning: A longitudinal study of Danish twins aged 70 years and older. *Behavior Genetics*, 33 (2), 125–136. https://doi.org/10.1023/A:1022501817781

Clapham, D.H., Dormling, I., Ekberg, I., Eriksson, G., Qamaruddin, M. & Vince-Prue (1998). Latitudinal cline of requirement for far-red light for the photoperiodic control of budset and extension growth in Picea abies (Norway spruce). *Physiologia Plantarum*. https://doi.org/10.1034/j.1399-3054.1998.1020110.x

Corbett-Detig, R.B., Zhou, J., Clark, A.G., Hartl, D.L. & Ayroles, J.F. (2013). Genetic incompatibilities are widespread within species. *Nature*, 504 (7478), 135–137. https://doi.org/10.1038/nature12678

Cronk, Q.C.B. (2005). Plant eco-devo: The potential of poplar as a model organism. *New Phytologist*, 166 (1), 39–48. https://doi.org/10.1111/j.1469-8137.2005.01369.x

Crow, J.F. (1948). Alternative Hypotheses of Hybrid Vigor. *Genetics*, 33 (5), 477–487. https://doi.org/10.1093/genetics/33.5.477

Dickmann DI, Kuzovkina J. Poplars and willows of the world, with emphasis on silviculturally important species. Poplars Willows Trees Soc Environ. 2014;22:8.

Do, R., Balick, D., Li, H., Adzhubei, I., Sunyaev, S. & Reich, D. (2015). No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nature Genetics*, 47 (2), 126–131. https://doi.org/10.1038/ng.3186

Doorenbos, J. (1953). Review of the literature on dormancy in buds of woody plants. *Mededelingen van de Landbouwhogeschool the Wageningen/Nederland*, 53 (1), 1–24

Duvick, D.N. (2001). Biotechnology in the 1930s: The development of hybrid maize. *Nature Reviews Genetics*, 2 (1), 69–74. https://doi.org/10.1038/35047587

Einspahr, D.W. (1984). Production and utilization of triploid hybrid aspen [Populus tremuloides, Populus tremula]. *Iowa State Journal of Research*

Elena, S.F. & Lenski, R.E. (1997). Test of synergistic interactions among deleterious mutations in bacteria. *Nature*, 390 (6658), 395–398. https://doi.org/10.1038/37108

Evans, L.M., Slavov, G.T., Rodgers-Melnick, E., Martin, J., Ranjan, P., Muchero, W., Brunner, A.M., Schackwitz, W., Gunter, L., Chen, J.-G., Tuskan, G. a

& DiFazio, S.P. (2014). Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nature Genetics*, advance on (10), 1089–1096. https://doi.org/10.1038/ng.3075

Fadista, J., Manning, A.K., Florez, J.C. & Groop, L. (2016). The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European Journal of Human Genetics*, 24 (8), 1202–1205. https://doi.org/10.1038/ejhg.2015.269

Falconer, D. S. (1981). *Introduction to Quantitative Genetics, 2nd Edn.* New York, NY: Longman Inc.

Felsenstein, J. (1974). The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics*, 78, 737–756

Flint-Garcia, S.A., Thornsberry, J.M. & Edward IV, S.B. (2003). Structure of Linkage Disequilibrium in Plants. *Annual Review of Plant Biology*, 54, 357–374. https://doi.org/10.1146/annurev.arplant.54.031902.134907

Fournier-Level, A., Korte, A., Cooper, M.D., Nordborg, M., Schmitt, J. & Wilczek, A.M. (2011). A map of local adaptation in Arabidopsis thaliana. *Science*, 334 (6052), 86–89. https://doi.org/10.1126/science.1209271

Fracheboud, Y., Luquez, V., Björkén, L., Sjödin, A., Tuominen, H. & Jansson, S. (2009). The control of autumn senescence in European aspen1[w][OA]. *Plant Physiology*, 149 (4), 1982–1991. https://doi.org/10.1104/pp.108.133249

Frewen, B.E., Chen, T.H.H., Howe, G.T., Davis, J., Antje Rohde, Boerjan, W. & Bradshaw, H.D. (2000). Quantitative trait loci and candidate gene mapping of bud set and bud flush in populus. *Genetics*, 154 (2), 837–845

Fullerton, S.M., Carvalho, A.B. & Clark, A.G. (2001). Letter to the Editor Local Rates of Recombination Are Positively Correlated with GC Content in the Human Genome. *Molecular Biology and Evolution*, 18 (6), 1139–1142

Gerstein, A.C. (2013). Mutational effects depend on ploidy level: all else is not equal. *Biology Letters*, 9 (1), 20120614. https://doi.org/10.1098/rsbl.2012.0614

Ghelardini, L., Berlin, S., Weih, M., Lagercrantz, U., Gyllenstrand, N. & Rönnberg-Wästljung, A.C. (2014). Genetic architecture of spring and autumn phenology in Salix. *BMC Plant Biology*, 14 (1), 1–18. https://doi.org/10.1186/1471-2229-14-31

Giraut, L., Falque, M., Drouaud, J., Pereira, L., Martin, O.C. & Mézard, C. (2011). Genome-wide crossover distribution in Arabidopsis thaliana meiosis reveals sex-specific patterns along chromosomes. *PLoS Genetics*, 7 (11). https://doi.org/10.1371/journal.pgen.1002354

Goodnight, C.J. (1999). Epistasis and Heterosis. *Genetics and Exploitation of Heterosis in Crops*. https://doi.org/https://doi.org/10.2134/1999.geneticsandexploitation.c6

Gougherty, A. V., Keller, S.R. & Fitzpatrick, M.C. (2021). Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. *Nature Climate Change*, 11 (2), 166–171. https://doi.org/10.1038/s41558-020-00968-6

Grattapaglia, D. (2004). Integrating genomics into Eucalyptus breeding. *Genetics and molecular research : GMR*, 3 (3), 369–379

Greene, D.F. & Johnson, E.A. (1999). Modelling recruitment of *Populus tremuloides*, *Pinus banksiana*, and *Picea mariana* following fire in the mixedwood boreal forest. *Canadian Journal of Forest Research*, 29 (4), 462–473. https://doi.org/10.1139/cjfr-29-4-462

Haldane, J.B. (2004). The rate of spontaneous mutation of a human gene. 1935. *Journal of genetics*, 83 (3), 235–244. https://doi.org/10.1007/BF02717892

Haldane, J.B.S. (1935). The rate of spontaneous mutation of a human gene. *Journal of Genetics*, 31 (3), 317–326. https://doi.org/10.1007/BF02982403

Hallmark, B., Karafet, T.M., Hsieh, P.H., Osipova, L.P., Watkins, J.C. & Hammer, M.F. (2019). Genomic evidence of local adaptation to climate and diet in indigenous Siberians. *Molecular Biology and Evolution*, 36 (2), 315–327. https://doi.org/10.1093/molbev/msy211

He, L. & Dooner, H.K. (2009). Haplotype structure strongly affects recombination in a maize genetic interval polymorphic for Helitron and retrotransposon insertions. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (21), 8410–8416. https://doi.org/10.1073/pnas.0902972106

Hemani, G., Knott, S. & Haley, C. (2013). An Evolutionary Perspective on Epistasis and the Missing Heritability. *PLoS Genetics*, 9 (2). https://doi.org/10.1371/journal.pgen.1003295

Henderson, I.R. (2012). Control of meiotic recombination frequency in plant genomes. *Current Opinion in Plant Biology*, 15 (5), 556–561. https://doi.org/10.1016/j.pbi.2012.09.002

Henn, B.M., Botigué, L.R., Bustamante, C.D., Clark, A.G. & Gravel, S. (2015). Estimating the mutation load in human genomes. *Nature Publishing Group*, 16 (6), 1–11. https://doi.org/10.1038/nrg3931

Henn, B.M., Botigué, L.R., Peischl, S., Dupanloup, I., Lipatov, M., Maples, B.K., Martin, A.R., Musharoff, S., Cann, H., Snyder, M.P., Excoffier, L., Kidd, J.M. & Bustamante, C.D. (2016). Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proceedings of the National Academy of Sciences*, 113 (4), E440–E449. https://doi.org/10.1073/pnas.1510805112

Hill, W.G. & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research*, 8 (3), 269–294. https://doi.org/10.1017/S0016672300010156

Hosius, B., Leinemann, L., Konnert, M. & Bergmann, F. (2006). Genetic aspects of forestry in the Central Europe. *European Journal of Forest Research*, 125 (4), 407–417. https://doi.org/10.1007/s10342-006-0136-4

Houben, A. & Schubert, I. (2003). DNA and proteins of plant centromeres. *Current Opinion in Plant Biology*, 6 (6), 554–560. https://doi.org/10.1016/j.pbi.2003.09.007

Howe, G.T., Aitken, S.N., Neale, D.B., Jermstad, K.D., Wheeler, N.C. & Chen, T.H.H. (2003). From genotype to phenotype: Unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany*, 81 (12), 1247–1266. https://doi.org/10.1139/b03-141

Hurme, P., Repo, T., Savolainen, O. & Paakkonen, T. (1997). Climatic adaptation of bud set and frost hardiness in Scots pine (Pinus sylvestris). *Canadian Journal of Forest Research*, 27 (5), 716–723. https://doi.org/10.1139/x97-052

Ingvarsson, P.K. (2010). Natural Selection on Synonymous and Nonsynonymous Mutations Shapes Patterns of Polymorphism in Populus tremula. *Molecular Biology and Evolution*, 27 (3), 650–660. https://doi.org/10.1093/molbev/msp255

Ingvarsson, P.K. & Bernhardsson, C. (2020). Genome-wide signatures of environmental adaptation in European aspen (Populus tremula) under current and future climate conditions. *Evolutionary Applications*, 13 (1), 132–142. https://doi.org/10.1111/eva.12792

Jansson, S. & Douglas, C.J. (2007). Populus : A Model System for Plant Biology . *Annual Review of Plant Biology*, 58 (1), 435–458. https://doi.org/10.1146/annurev.arplant.58.032806.103956

Josephs, E.B., Stinchcombe, J.R. & Wright, S.I. (2017). What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytologist*, 214 (1), 21–33. https://doi.org/10.1111/nph.14410

Kalcsits, L.A., Silim, S. & Tanino, K. (2009). Warm temperature accelerates short photoperiod-induced growth cessation and dormancy induction in hybrid poplar (populus × spp.). *Trees - Structure and Function*, 23 (5), 971–979. https://doi.org/10.1007/s00468-009-0339-7

Karacic, A., Verwijst, T. & Weih, M. (2003). Above-ground woody biomass production of short-rotation Populus plantations on agricultural land in Sweden. *Scandinavian Journal of Forest Research*, 18 (5), 427–437. https://doi.org/10.1080/02827580310009113

Kawecki, T.J. & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, 7 (12), 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Keenan, T.F. & Richardson, A.D. (2015). The timing of autumn senescence is affected by the timing of spring phenology: Implications for predictive

models. *Global Change Biology*, 21 (7), 2634–2641. https://doi.org/10.1111/gcb.12890

Keskitalo, J., Bergquist, G., Gardeström, P. & Jansson, S. (2005). A cellular timetable of autumn senescence. *Plant Physiology*, 139 (4), 1635–1648. https://doi.org/10.1104/pp.105.066845

Kim, S., Plagnol, V., Hu, T.T., Toomajian, C., Clark, R.M., Ossowski, S., Ecker, J.R., Weigel, D. & Nordborg, M. (2007). Recombination and linkage disequilibrium in Arabidopsis thaliana. *Nature Genetics*, 39 (9), 1151–1155. https://doi.org/10.1038/ng2115

Kono, T.J.Y., Lei, L., Shih, C., Hoffman, P.J., Morrell, P.L. & Fay, J.C. (2018). Comparative Genomics Approaches Accurately Predict Deleterious Variants in Plants. *G3: Genes|Genomes|Genetics*, 8 (October), 3321–3329. https://doi.org/10.1534/g3.118.200563

Kouki, J., Arnold, K. & Martikainen, P. (2004). Long-term persistence of aspen – a key host for many threatened species – is endangered in old-growth conservation areas in Finland. *Journal for Nature Conservation*, 12, 41–52. https://doi.org/10.1016/j.jnc.2003.08.002

Kulathinal, R.J., Bennett, S.M., Fitzpatrick, C.L. & Noor, M.A.F. (2008). Fine-scale mapping of recombination rate in Drosophila refines its correlation to diversity and divergence. *Proceedings of the National Academy of Sciences*, 105 (29), 10051–10056. https://doi.org/10.1073/pnas.0801848105

Kumar, P., Henikoff, S. & Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4 (8), 1073–1081. https://doi.org/10.1038/nprot.2009.86

Lang, G.I., Rice, D.P., Hickman, M.J., Sodergren, E., Weinstock, G.M., Botstein, D. & Desai, M.M. (2013). Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, 500 (7464), 571–574. https://doi.org/10.1038/nature12344

Larson, G., Piperno, D.R., Allaby, R.G., Purugganan, M.D., Andersson, L., Arroyo-Kalin, M., Barton, L., Vigueira, C.C., Denham, T., Dobney, K., Doust, A.N., Gepts, P., Gilbert, M.T.P., Gremillion, K.J., Lucas, L., Lukens, L., Marshall, F.B., Olsen, K.M., Pires, J.C., Richerson, P.J., De Casas, R.R., Sanjur, O.I., Thomas, M.G. & Fuller, D.Q. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences of the United States of America*, 111 (17), 6139–6146. https://doi.org/10.1073/pnas.1323964111

Leimu, R. & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS ONE*, 3 (12), 1–8. https://doi.org/10.1371/journal.pone.0004010

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. (2014). UpSet: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, 20 (12), 1983–1992. https://doi.org/10.1109/TVCG.2014.2346248

Li, B., McKeand, S. & Weir, R. (1999). Tree improvement and sustainable forestry–impact of two cycles of loblolly pine breeding in the USA. *Forest Genetics*, 6 (4), 229–234. http://www.tuzvo.sk/files/fg/volumes/1999/FG06-4_229_234.pdf

Li, Z.K., Luo, L.J., Mei, H.W., Wang, D.L., Shu, Q.Y., Tabien, R., Zhong, D.B., Ying, C.S., Stansel, J.W., Khush, G.S. & Paterson, A.H. (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics*, 158 (4), 1737–1753. https://doi.org/10.1093/genetics/158.4.1737

Lin, Y.-C., Wang, J., Delhomme, N., Schiffthaler, B., Sundström, G., Zuccolo, A., Nystedt, B., Hvidsten, T.R., de la Torre, A., Cossu, R.M., Hoeppner, M.P., Lantz, H., Scofield, D.G., Zamani, N., Johansson, A., Mannapperuma, C., Robinson, K.M., Mähler, N., Leitch, I.J., Pellicer, J., Park, E.-J., Van Montagu, M., Van de Peer, Y., Grabherr, M., Jansson, S., Ingvarsson, P.K. & Street, N.R. (2018). Functional and evolutionary genomic inferences in *Populus* through genome and population sequencing of American and European aspen. *Proceedings of the National Academy of Sciences*, 201801437. https://doi.org/10.1073/pnas.1801437115

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

Ljung, K., Bhalerao, R.P. & Sandberg, G. (2001). Sites and homeostatic control of auxin biosynthesis in Arabidopsis during vegetative growth. *Plant Journal*, 28 (4), 465–474. https://doi.org/10.1046/j.1365-313X.2001.01173.x

Loehle, C. (1998). Height growth rate tradeoffs determine northern and southern range limits for trees. *Journal of Biogeography*, 25, 735–742

Lucek, K. & Willi, Y. (2021). Drivers of linkage disequilibrium across a species' geographic range. *PLoS Genetics*, 17 (3), 1–22. https://doi.org/10.1371/journal.pgen.1009477

Luquez, V., Hall, D., Albrectsen, B.R., Karlsson, J., Ingvarsson, P. & Jansson, S. (2008). Natural phenological variation in aspen (Populus tremula): the SwAsp collection. *Tree Genetics & Genomes*, 4 (2), 279–292. https://doi.org/10.1007/s11295-007-0108-y

Lynch, D.B., Logue, M.E., Butler, G. & Wolfe, K.H. (2010). Chromosomal G + C content evolution in yeasts: Systematic interspecies differences, and GC-poor troughs at centromeres. *Genome Biology and Evolution*, 2 (1), 572–583. https://doi.org/10.1093/gbe/evq042

Lynch, M. (1991). The Genetic Interpretation of Inbreeding Depression and Outbreeding Depression. *Evolution*, 45 (3), 622. https://doi.org/10.2307/2409915

Lynch, M., Butcher, D., Bürger, R. & Gabriel, W. (1993). The mutational meltdown in asexual populations. *Journal of Heredity*, 84 (5), 339–344. https://doi.org/10.1093/oxfordjournals.jhered.a111354

Ma, T., Wang, J., Zhou, G., Yue, Z., Hu, Q., Chen, Y., Liu, B., Qiu, Q., Wang, Z., Zhang, J., Wang, K., Jiang, D., Gou, C., Yu, L., Zhan, D., Zhou, R., Luo, W., Ma, H., Yang, Y., Pan, S., Fang, D., Luo, Y., Wang, X., Wang, G., Wang, J., Wang, Q., Lu, X., Chen, Z., Liu, J., Lu, Y., Yin, Y., Yang, H., Abbott, R.J., Wu, Y., Wan, D., Li, J., Yin, T., Lascoux, M., DiFazio, S.P., Tuskan, G.A., Wang, J. & Liu, J. (2013). Genomic insights into salt adaptation in a desert poplar. *Nature Communications*, 4 (1), 2797. https://doi.org/10.1038/ncomms3797

Mancera, E., Bourgon, R., Brozzi, A., Huber, W. & Steinmetz, L.M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature*, 454 (7203), 479–485. https://doi.org/10.1038/nature07135

Mandel, M. & Marmur, J. (1968). [109] Use of ultraviolet absorbance-temperature profile for determining the guanine plus cytosine content of DNA. *Methods in Enzymology*, 12, 195–206

Marand, A.P., Zhao, H., Zhang, W., Zeng, Z., Fang, C. & Jianga, J. (2019). Historical meiotic crossover hotspots fueled patterns of evolutionary divergence in rice. *Plant Cell*, 31 (3), 645–662. https://doi.org/10.1105/tpc.18.00750

Marron, N., Storme, V., Dillen, S.Y., Bastien, C., Ricciotti, L., Salani, F., Sabatti, M., Rae, A.M., Ceulemans, R. & Boerjan, W. (2010). Genomic regions involved in productivity of two interspecific poplar families in Europe. 2. Biomass production and its relationships with tree architecture and phenology. *Tree Genetics and Genomes*, 6 (4), 533–554. https://doi.org/10.1007/s11295-010-0270-5

Marsden, C.D., Ortega-Del Vecchyo, D., O'Brien, D.P., Taylor, J.F., Ramirez, O., Vilà, C., Marques-Bonet, T., Schnabel, R.D., Wayne, R.K. & Lohmueller, K.E. (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proceedings of the National Academy of Sciences*, 113 (1), 152–157. https://doi.org/10.1073/pnas.1512501113

McKown, A.D., Guy, R.D., Klápště, J., Geraldes, A., Friedmann, M., Cronk, Q.C.B., El-Kassaby, Y.A., Mansfield, S.D. & Douglas, C.J. (2014a). Geographical and environmental gradients shape phenotypic trait variation and genetic structure in P opulus trichocarpa. *New Phytologist*, 201 (4), 1263–1276. https://doi.org/10.1111/nph.12601

McKown, A.D., Klápště, J., Guy, R.D., Geraldes, A., Porth, I., Hannemann, J., Friedmann, M., Muchero, W., Tuskan, G.A., Ehlting, J., Cronk, Q.C.B., El-Kassaby, Y.A., Mansfield, S.D. & Douglas, C.J. (2014b). Genome-wide association implicates numerous genes underlying ecological trait variation

in natural populations of Populus trichocarpa. *New Phytologist*, 203 (2), 535–553. https://doi.org/10.1111/nph.12815

McVean, G.A.T. (2004). The Fine-Scale Structure of Recombination Rate Variation in the Human Genome. *Science*, 304 (5670), 581–584. https://doi.org/10.1126/science.1092500

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157 (4), 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Moran, G.F., Bell, J.G. & Matheson, A.C. (1980). The genetic structure and levels of inbreeding in a Pinus radiata D. Don seed orchard. *Silvae Genetica*, 29 (5/6), 190–193. https://www.cabdirect.org/cabdirect/abstract/19810672366

Morgan, T.H. (1910). Sex limited inheritance in drosophila. *Science*, 32 (812), 120–122. https://doi.org/10.1126/science.32.812.120

Morgan, T.H. (1915). The mechanism of Mendelian heredity. Henry Holt and Company,: New York :

Moyers, B.T., Morrell, P.L. & McKay, J.K. (2018). Genetic Costs of Domestication and Improvement. *Journal of Heredity*, 109 (2), 103–116. https://doi.org/10.1093/jhered/esx069

Munné-Bosch, S. & Alegre, L. (2004). Die and let live: Leaf senescence contributes to plant survival under drought stress. *Functional Plant Biology*, 31 (3), 203–216. https://doi.org/10.1071/FP03236

Nachman, M. (2002). Variation in recombination rate across the genome: evidence and implications. *Current Opinion in Genetics & Development*, 12 (6), 657–663. https://doi.org/10.1016/S0959-437X(02)00358-1

Nicotra, A.B., Atkin, O.K., Bonser, S.P., Davidson, A.M., Finnegan, E.J., Mathesius, U., Poot, P., Purugganan, M.D., Richards, C.L., Valladares, F. & van Kleunen, M. (2010). Plant phenotypic plasticity in a changing climate. *Trends in Plant Science*, 15 (12), 684–692. https://doi.org/10.1016/j.tplants.2010.09.008

Nordborg, M. & Weigel, D. (2008). Next-generation genetics in plants. *Nature*, 456 (7223), 720–723. https://doi.org/10.1038/nature07629

Orr-Weaver, T.L. (2015). When bigger is better: The role of polyploidy in organogenesis. *Trends in Genetics*, 31 (6), 307–315. https://doi.org/10.1016/j.tig.2015.03.011

Paolucci, I., Gaudet, M., Jorge, V., Beritognolo, I., Terzoli, S., Kuzminsky, E., Muleo, R., Mugnozza, G.S. & Sabatti, M. (2010). Genetic linkage maps of Populus alba L. and comparative mapping analysis of sex determination across Populus species. *Tree Genetics and Genomes*, 6 (6), 863–875. https://doi.org/10.1007/s11295-010-0297-7

Pearson, T.A. & Manolio, T.A. (2008). How to interpret a genome-wide association study. *JAMA - Journal of the American Medical Association*, 299 (11), 1335–1344. https://doi.org/10.1001/jama.299.11.1335

Pliura, A., Suchockas, V., Sarsekova, D. & Gudynaite, V. (2014). Genotypic variation and heritability of growth and adaptive traits, and adaptation of young poplar hybrids at northern margins of natural distribution of Populus nigra in Europe. *Biomass and Bioenergy*, 70, 513–529. https://doi.org/10.1016/j.biombioe.2014.09.011

Qiu, Y., Guan, S.C., Wen, C., Li, P., Gao, Z. & Chen, X. (2019). Auxin and cytokinin coordinate the dormancy and outgrowth of axillary bud in strawberry runner. *BMC Plant Biology*, 19 (1), 1–16. https://doi.org/10.1186/s12870-019-2151-x

Ramu, P., Esuma, W., Kawuki, R., Rabbi, I.Y., Egesi, C., Bredeson, J. V, Bart, R.S., Verma, J., Buckler, E.S. & Lu, F. (2017). Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nature Genetics*, 49 (March), 959–963. https://doi.org/10.1038/ng.3845

Raymond, O., Gouzy, J., Just, J., Badouin, H., Verdenaud, M., Lemainque, A., Vergne, P., Moja, S., Choisne, N., Pont, C., Carrère, S., Caissard, J.C., Couloux, A., Cottret, L., Aury, J.M., Szécsi, J., Latrasse, D., Madoui, M.A., François, L., Fu, X., Yang, S.H., Dubois, A., Piola, F., Larrieu, A., Perez, M., Labadie, K., Perrier, L., Govetto, B., Labrousse, Y., Villand, P., Bardoux, C., Boltz, V., Lopez-Roques, C., Heitzler, P., Vernoux, T., Vandenbussche, M., Quesneville, H., Boualem, A., Bendahmane, A., Liu, C., Le Bris, M., Salse, J., Baudino, S., Benhamed, M., Wincker, P. & Bendahmane, M. (2018). The Rosa genome provides new insights into the domestication of modern roses. *Nature Genetics*, 50 (6), 772–777. https://doi.org/10.1038/s41588-018-0110-3

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution*, 66 (1), 1–17. https://doi.org/10.1111/j.1558-5646.2011.01486.x

Rodgers-Melnick, E., Bradbury, P.J., Elshire, R.J., Glaubitz, J.C., Acharya, C.B., Mitchell, S.E., Li, C., Li, Y. & Buckler, E.S. (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences of the United States of America*, 112 (12), 3823–8. https://doi.org/10.1073/pnas.1413864112

Rohde, A., Storme, V., Jorge, V., Gaudet, M., Vitacolonna, N., Fabbrini, F., Ruttink, T., Zaina, G., Marron, N., Dillen, S., Steenackers, M., Sabatti, M., Morgante, M., Boerjan, W. & Bastien, C. (2011). Bud set in poplar - genetic dissection of a complex trait in natural and hybrid populations. *New Phytologist*, 189 (1), 106–121. https://doi.org/10.1111/j.1469-8137.2010.03469.x

Rönnberg-Wästljung, A.C. & Gullberg, U. (1999). Genetics of breeding characters with possible effects on biomass production in Salix viminalis (L.). *Theoretical and Applied Genetics*, 98 (3–4), 531–540. https://doi.org/10.1007/s001220051101

Sanjuan, R. & Nebot, M.R. (2008). A Network Model for the Correlation between Epistasis and Genomic Complexity. *PloS one*, 3 (7), e2663. https://doi.org/doi:10.1371/ journal.pone.0002663

Sannigrahi, P., Ragauskas, A.J. & Tuskan, G.A. (2010). Poplar as a feedstock for biofuels: A review of compositional characteristics. *Biofuels, Bioproducts and Biorefining*, 4 (2), 209–226. https://doi.org/10.1002/bbb.206

Savolainen, O., Pyhäjärvi, T. & Knürr, T. (2007). Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics*, 38, 595–619. https://doi.org/10.1146/annurev.ecolsys.38.091206.095646

Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., Albrechtsen, A., Dupanloup, I., Foucal, A., Petersen, B., Fumagalli, M., Raghavan, M., Seguin-Orlando, A., Korneliussen, T.S., Velazquez, A.M. V, Stenderup, J., Hoover, C.A., Rubin, C.-J., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., MacHugh, D.E., Kalbfleisch, T., MacLeod, J.N., Rubin, E.M., Sicheritz-Ponten, T., Andersson, L., Hofreiter, M., Marques-Bonet, T., Gilbert, M.T.P., Nielsen, R., Excoffier, L., Willerslev, E., Shapiro, B. & Orlando, L. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences*, 111 (52), E5661–E5669. https://doi.org/10.1073/pnas.1416991111

Serba, D.D., Muleta, K.T., St. Amand, P., Bernardo, A., Bai, G., Perumal, R. & Bashir, E. (2019). Genetic Diversity, Population Structure, and Linkage Disequilibrium of Pearl Millet. *The Plant Genome*, 12 (3), 180091. https://doi.org/10.3835/plantgenome2018.11.0091

Shanfelter, A.F., Archambeault, S.L. & White, M.A. (2019). Divergent Fine-Scale Recombination Landscapes between a Freshwater and Marine Population of Threespine Stickleback Fish. *Genome Biology and Evolution*, 11 (6), 1573–1585. https://doi.org/10.1093/gbe/evz090

Shull, G.H. (1908). The Composition of a Field of Maize. *Journal of Heredity*, os-4 (1), 296–301. https://doi.org/10.1093/jhered/os-4.1.296

Si, W., Yuan, Y., Huang, J., Zhang, X., Zhang, Y., Zhang, Y., Tian, D., Wang, C., Yang, Y. & Yang, S. (2015). Widely distributed hot and cold spots in meiotic recombination as shown by the sequencing of rice F2 plants. *New Phytologist*, 206 (4), 1491–1502. https://doi.org/10.1111/nph.13319

Silva-Junior, O.B. & Grattapaglia, D. (2015). Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of E ucalyptus grandis. *New Phytologist*, 208 (3), 830–845. https://doi.org/10.1111/nph.13505

Singh, R.K., Maurya, J.P., Azeez, A., Miskolczi, P., Tylewicz, S., Stojkovič, K., Delhomme, N., Busov, V. & Bhalerao, R.P. (2018). A genetic network

mediating the control of bud break in hybrid aspen. *Nature Communications*, 9 (1). https://doi.org/10.1038/s41467-018-06696-y

Slatkin, M. (2008). Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9 (6), 477–485. https://doi.org/10.1038/nrg2361

Slavov, G.T., DiFazio, S.P., Martin, J., Schackwitz, W., Muchero, W., Rodgers-Melnick, E., Lipphardt, M.F., Pennacchio, C.P., Hellsten, U., Pennacchio, L.A., Gunter, L.E., Ranjan, P., Vining, K., Pomraning, K.R., Wilhelm, L.J., Pellegrini, M., Mockler, T.C., Freitag, M., Geraldes, A., El-Kassaby, Y.A., Mansfield, S.D., Cronk, Q.C.B., Douglas, C.J., Strauss, S.H., Rokhsar, D. & Tuskan, G.A. (2012). Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree Populus trichocarpa. *New Phytologist*, 196 (3), 713–725. https://doi.org/10.1111/j.1469-8137.2012.04258.x

Smith, A. V., Thomas, D.J., Munro, H.M. & Abecasis, G.R. (2005). Sequence features in regions of weak and strong linkage disequilibrium. *Genome Research*, 15 (11), 1519–1534. https://doi.org/10.1101/gr.4421405

Smith, J.M. & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, 23 (1), 23–35. https://doi.org/10.1017/S0016672300014634

Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W. & Smadja, C.M. (2017). Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372 (1736), 20160455. https://doi.org/10.1098/rstb.2016.0455

Suontama, M., Klápště, J., Telfer, E., Graham, N., Stovold, T., Low, C., McKinley, R. & Dungey, H. (2019). Efficiency of genomic prediction across two Eucalyptus nitens seed orchards with different selection histories. *Heredity*, 122 (3), 370–379. https://doi.org/10.1038/s41437-018-0119-5

Tan, B., Grattapaglia, D., Martins, G.S., Ferreira, K.Z., Sundberg, B. & Ingvarsson, P.K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two Eucalyptus species and their F1 hybrids. *BMC Plant Biology*, 17 (1), 1–15. https://doi.org/10.1186/s12870-017-1059-6

Tansey, C.J., Hadfield, J.D. & Phillimore, A.B. (2017). Estimating the ability of plants to plastically track temperature-mediated shifts in the spring phenological optimum. *Global Change Biology*, 23 (8), 3321–3334. https://doi.org/10.1111/gcb.13624

Taylor, G. (2002). Populus: Arabidopsis for Forestry. Do We Need a Model Tree? *Annals of Botany*, 90 (6), 681–689. https://doi.org/10.1093/aob/mcf255

Tiley, G.P. & Burleigh, G. (2015). The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. *BMC*

Evolutionary Biology, 15 (1), 1–14. https://doi.org/10.1186/s12862-015-0473-3

Tong, C., Li, H., Wang, Y., Li, X., Ou, J., Wang, D., Xu, H., Ma, C., Lang, X., Liu, G., Zhang, B. & Shi, J. (2016). Construction of high-density linkage maps of Populus deltoides × P. simonii using restriction-site associated DNA sequencing. *PLoS ONE*, 11 (3), 1–20. https://doi.org/10.1371/journal.pone.0150692

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.-L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Déjardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjärvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leplé, J.-C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouzé, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.-J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y. & Rokhsar, D. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science (New York, N.Y.)*, 313 (5793), 1596–604. https://doi.org/10.1126/science.1128691

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A. & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, 101 (1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Wang, J., Ding, J., Tan, B., Robinson, K.M., Michelson, I.H., Johansson, A., Nystedt, B., Scofield, D.G., Nilsson, O., Jansson, S., Street, N.R. & Ingvarsson, P.K. (2018). A major locus controls local adaptation and adaptive life history variation in a perennial plant. *Genome Biology*, 19 (1), 1–17. https://doi.org/10.1186/s13059-018-1444-y

Wang, J., Street, N.R., Scofield, D.G. & Ingvarsson, P.K. (2016). Natural selection and recombination rate variation shape nucleotide polymorphism across the genomes of three related populus species. *Genetics*, 202 (3), 1185–1200. https://doi.org/10.1534/genetics.115.183152

Waters, C.D., Hard, J.J., Brieuc, M.S.O., Fast, D.E., Warheit, K.I., Knudsen, C.M., Bosch, W.J. & Naish, K.A. (2018). Genomewide association analyses of fitness traits in captive-reared Chinook salmon: Applications in evaluating conservation strategies. *Evolutionary Applications*, 11 (6), 853–868. https://doi.org/10.1111/eva.12599

Weighill, D., Macaya-Sanz, D., DiFazio, S.P., Joubert, W., Shah, M., Schmutz, J., Sreedasyam, A., Tuskan, G. & Jacobson, D. (2019). Wavelet-based genomic signal processing for centromere identification and hypothesis generation. *Frontiers in Genetics*, 10 (MAY). https://doi.org/10.3389/fgene.2019.00487

Weih, M. (2004). Intensive short rotation forestry in boreal climates: Present and future perspectives. *Canadian Journal of Forest Research*, 34 (7), 1369–1378. https://doi.org/10.1139/X04-090

Willi, Y., Fracassetti, M., Zoller, S. & Van Buskirk, J. (2018). Accumulation of mutational load at the edges of a species range. *Molecular Biology and Evolution*, 35 (4), 781–791. https://doi.org/10.1093/molbev/msy003

Wilson, A.J., Kruuk, L.E.B. & Coltman, D.W. (2005). Ontogenetic patterns in heritable variation for body size: using random regression models in a wild ungulate population. *The American naturalist*, 166 (6). https://doi.org/10.1086/497441

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16, 97–159

Xie, Y., Wang, X., Wilson, A.M. & Silander, J.A. (2018). Predicting autumn phenology: How deciduous tree species respond to weather stressors. *Agricultural and Forest Meteorology*, 250–251 (June 2017), 127–137. https://doi.org/10.1016/j.agrformet.2017.12.259

Ye, S., Gao, N., Zheng, R., Chen, Z.T., Teng, J., Yuan, X., Zhang, H., Chen, Z., Zhang, X., Li, J. & Zhang, Z. (2019). Strategies for obtaining and pruning imputed whole-genome sequence data for genomic prediction. *Frontiers in Genetics*, 10 (JUN), 1–11. https://doi.org/10.3389/fgene.2019.00673

Yin, L., Zhang, H., Zhou, X., Yuan, X., Zhao, S., Li, X. & Liu, X. (2020). KAML: Improving genomic prediction accuracy of complex traits using machine learning determined parameters. *Genome Biology*, 21 (1), 1–22. https://doi.org/10.1186/s13059-020-02052-w

Yu, Q., Tigerstedt, P.M.A. & Haapanen, M. (2001). Growth and phenology hybrid aspen clones (Populus tremula L.X Populus Tremuloides Michx.). *Silva Fennica*, 35 (1), 15–25. https://doi.org/10.14214/sf.600

Zeder, M.A. (2011). The origins of agriculture in the Near East. *Current Anthropology*, 52 (SUPPL. 4). https://doi.org/10.1086/659307

Zhang, D.Z.Z. & Li, K.Y.B. (2004). Genetic mapping in (Populus tomentosa ” Populus bolleana) and P . tomentosa Carr . using AFLP markers. 657–662. https://doi.org/10.1007/s00122-003-1478-7

Zhang, M., Suren, H. & Holliday, J.A. (2019). Phenotypic and Genomic Local Adaptation across Latitude and Altitude in Populus trichocarpa. *Genome biology and evolution*, 11 (8), 2256–2272. https://doi.org/10.1093/gbe/evz151

Zhang, M., Zhou, L., Bawa, R., Suren, H. & Holliday, J.A. (2016). Recombination Rate Variation, Hitchhiking, and Demographic History Shape Deleterious Load in Poplar. *Molecular Biology and Evolution*, 33 (11), 2899–2910. https://doi.org/10.1093/molbev/msw169

Zhigunov, A. V, Ulianich, P.S., Lebedeva, M. V, Chang, P.L., Nuzhdin, S. V & Potokina, E.K. (2017). Development of F1 hybrid population and the high-density linkage map for European aspen (Populus tremula L .) using RADseq technology. 17 (Suppl 1). https://doi.org/10.1186/s12870-017-1127-y

Popular science summary

Interest in poplar and aspen species in the conifer dominated Swedish forestry sector arose in the beginning of 1990's, when landowners started utilizing unused agricultural land for production of bioenergy. Poplars and aspens are especially suited for production of bioenergy due to the short generation time of about 20 years and lack of need for replanting as most poplar and aspen species reproduce clonally through root suckers, making stands of poplars relatively easy to upkeep. One of the poplar species garnering interest is black cottonwood, a species native to North America, but the currently available commercial variants are ill fitting for the Swedish climate conditions, namely the relatively harsh winter compared to the native range of the species. Furthermore, the large physical size and consequent high space requirements of trees such as black cottonwood has limited attempts of breeding to rectify the poor fit to Swedish conditions, as breeding often requires between hundreds and thousands of individuals to be effective. In the work presented in this thesis, we desired to explore the opportunities of small population breeding in black cottonwood and the possibility of producing variants of the species suited to Swedish conditions to be used in bioenergy production using said population.

Our results showed that a breeding population of 121 black cottonwood individuals, mostly consisting of the offspring of only 19 parent individuals and as such being genetically somewhat limited, contained the required genetic resources to facilitate breeding for varieties better fitting to Swedish conditions. Furthermore, we located a handful of possible genetic variants to select for during breeding and proved that breeding work is required for adjusting the climate fit of black cottonwood, as novel combinations of genetic alleles were found to affect the before and after winter behavior of the trees in Sweden compared to the native range. Finally, we showed that

avoidance of careless breeding, leading to fixation of harmful genetic variants as seen in dog breeding in form of developmental issues and disease susceptibilities, can increase the breeding gain without need for hybridization, which is often utilized to produce top varieties in crop breeding. As such, our work in this thesis has shown that there are opportunities in small population breeding if the population has been produced from diverse enough parentage and if harmful genetic variants are kept an eye on during breeding.

Populärvetenskaplig sammanfattning

Intresset för poppel- och asparter som alternativ i det barrträdsdominerade svenska skogsbruket uppstod i början av 1990-talet, då markägare började använda outnyttjad jordbruksmark för produktion av bioenergi. Popplar och aspar är särskilt lämpade för produktion av bioenergi på grund av den korta rotationstiden på cirka 20 år och det faktum att de inte behöver omplanteras efter skörd då de flesta poppel- och asparter förökar sig genom klontillväxt vilket gör bestånden lätt att underhålla. En av poppelarterna som väckt störst intresse är jättepoppel, en art som ursprungligen kommer från Nordamerika. De kommersiella varianterna som för närvarande finns tillgängliga är dock inte väl anpassade till svenska klimatförhållandena, framfört allt den relativt långa och kalla vintern. Vidare leder den stora fysiska storleken och därmed det höga platsbehov dom träden kräver till att jättepoppel ofta odlas i begränsade försök. Detta leder i sin tur till att det kan vara svårt att bedriva förädling för att gynna anpassningen svenska förhållanden, eftersom förädling traditionellt kräver hundratals eller tusentals individer för att vara effektiv. I den här avhandlingen har vi varit intresserade av att utforska möjligheterna för hur man kan bedriva förädling av jättepoppel även med en begränsad förädlingspopulation och på så sätt producera sorter som passar svenska förhållanden för att kunna användas för bioenergiproduktion.

Våra resultat visar att en förädlingspopulation på 121 jättepoppelindivider, som mestadels har sitt ursprung från 19 föräldragenotyper, och som därför är att betrakta som genetiskt begränsad, trots detta innehöll de tillräckligt med genetisk variation för att möjliggöra en anpassning till svenska klimatförhållanden. Dessutom identifierade vi ett antal genetiska varianter som kan utnyttjas under förädlingen för att klimatanpassa jättepoppel till svenska förhållanden. Slutligen visar vi att man genom att systematiskt undvika fixering av skadliga mutationer så kan

man kan öka förädlingsvinsten utan att behöva förlita sig på hybridisering, som annas ofta används för att producera elitsorter hos olika grödor. Som sådan har arbetet som presenteras i denna avhandling visat att det är möjligt att arbeta med förädling även i små bestånd om populationen innehåller tillräckliga mängder genetisk variation och om skadliga genetiska varianter kan hållas i schack under förädlingsprocessen.

Acknowledgements

This thesis would not have become reality if not for the influence of many people. While only a limited number of people worked on this thesis physically, many more inspired or helped me in other ways during these four years. In this section I want to thank you all. I will start with by thanking my many colleagues and co-authors from both CLAP, Martin Weih, Giulia Vico, Almir Karacic, Anneli Adler and Magnus Hertzberg, and UmAsp/SwAsp context, Nathaniel Street, Bastian Schiffhaler, Kathryn Robinson and Stefan Jansson, for their contributions towards the papers constituting this thesis. I also want to thank the external evaluators of my thesis, Henrik Böhlenius, Jenny Hagenblad and Martin Lascoux, as well as the evaluators of my PhD progress across these years, Anna Westerbergh and Adrien Sicard for helping to make sure I am at the level required to be granted the esteemed title of PhD. I want to extend large thank you to Jason Holliday for acting as my opponent for my dissertation despite the massive time difference. I want to thank the administrative staff from the department of plant biology, Lotta Olsson, Qinq Heijkenskjöld, Björn Nicander and Christian Divander for all their help in making my PhD move smoothly. I also want to thank Carina Grönholm and Jacob Hoffman for giving me a comfortable place to live when I was facing the shadiness of Uppsala rental apartment market. I also want to thank my girlfriend Suzana for dealing with me in the lows and sharing with me the highs that this journey to PhD has contained.

I will do the rest of these acknowledgements in a roughly chronological order. I first and foremost want to thank my mother and father. They both inspired and accommodated my young self's interest in nature. Whether it be taking me bird watching or fishing, allowing me to join in the farm work or reading me animal encyclopaedia's for good night stories, they always gave me their all. Of course, it was not just during those early years, but I owe my

whole life to my parents and I owe so much of my personality, work ethics and well-being to them as well. I want to also thank my big brother Riku and big sister Riikka for being there for me all my life. Finally, I want to thank the Finnish Yle TV1 channel for their long running Saturday night nature documentary series “Avara luonto”, and my parents for letting me watch them from such a young age despite the somewhat gory way nature works sometimes, as they allowed me to see the diversity of the natural world and dream of being one day a brave biologist exploring nature.

One cannot really become an academic, unless they like, or at the very least tolerate, lectures and classes. Education and educating myself has been a large part of my life, now culminating in my PhD. As such, I want to thank my teachers Ritva Mäntymäki, Lassi Suni and Päivi Aarnio-Laukkanen for teaching me and showing that education and the subject of biology can be very rewarding and enjoyable. I also want to say that Päivi’s enthusiasm and love for her subject of teaching as my high school biology teacher was a considerable part of the reason I decided I wanted to be a biologist when I grow up, so thank you for that as well. Furthermore, I want to thank all my university professors, but especially professors Anja Hohtola and Outi Savolainen for teaching me much of what was used as the base for my PhD across my Batchelor’s and Master’s degrees and doing so in very memorable fashion both in their own ways. Your teachings have been invaluable to me. I want to thank Joanne Russell and Allan Booth from James Hutton institute in Dundee for giving me the first tastes of scientific work and the international nature of biology.

Professor Savolainen was not just my professor and teacher, but she was also the supervisor of my Master’s thesis work. During this time, she taught me a lot of what it means to do science. I cannot say all of it was pleasant, but life is not always pleasant, and it helped me to grow thicker skin. I want to thank everyone from the Savolainen group at that time, but I want to give special thanks to Sonja Kujala, who was one of the kindest and calmest people I have ever had the pleasure of meeting. Your softer touch on things gave my younger and meeker self hope that I could be a scientist. I also want to thank my Master’s degree co-supervisor Katri Kärkkäinen from Luke for the opportunity to work with Luke Scots pine field trials in both Muhos and Loppi. I also want to thank the Luke staff in both Oulu and Loppi for helping me collect the data I needed. Finally, I want to thank that kind soul who allowed a stranger into their home on Sunday night to call and wait for

someone to come pull his car out of the slushy road he in his stupidity drove the car into. I wish I remembered your name, but sadly I have forgotten, but the memory of your kind deed lives on.

Now we come to my time in Ingvarsson group working on my PhD and my time at SLU, and we come to a long list of people I want to thank. I want to thank Le Yu for her helpful comments to my thesis. I want to thank Carolina Bernhardsson and Biyue Tan for all of their help especially during the early days of my PhD, both being willing to help me almost at the drop of a hat. I want to thank Mathilde Bettembourg, Silvana Moreno, Biyue Tan and Xi Wang who all helped me to feel at home at SLU. Xi Wang I want to especially thank for those long Friday afternoon chats, where the cool parts of biology were often discussed and Mathilde Bettembourg for the random corridor chats that helped me to take my mind off work at many crucial times. I want to thank Girma Bedada for his insights to many of my presentations and to this thesis, and for pointing out things not thought of by others and thus help me produce more well-rounded science. I also want to thank Thomas Richards and Martha Rendón for all of their help both in science and outside of it. Their willingness to help whenever needed was always appreciated by me, but especially the shared experiences were valuable help for someone who quite often doubts himself. It was very reassuring to hear that the struggles and doubts I often faced were also experienced by such accomplished and skilled people in their days as PhD students. I also want to thank Tom for pitching me the CLAP -ball and helping me get up to speed with it. I want to thank my PhD co-supervisor Ann-Christin Rönnberg-Wästljung for always giving me time and fresh perspectives when I came looking for them. Finally, I want to thank professor Pär Ingvarsson for everything from offering to hire me just days before Christmas to this beautiful PhD thesis and everything in between. Pelle has been the perfect supervisor for me, allowing me to work independently, but always having his door open for when I needed help. Never pressuring me too much, but never letting me slack off either. Without Pelle, this thesis would never have become reality.

This thesis is thus not just for me, but all of you. It is a testament to the importance of surrounding yourself with the right kind of people, asking for help and working hard. No human is an island, and this is especially true for science. I want to thank once more everyone who have put effort to this thesis, I would not be here without you. Thank you!

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2021:60

Small populations often exhibit high levels of linkage disequilibrium (LD) and low levels of genetic diversity. This makes them poorly suited for breeding, where high LD can lead to numerous problems including mutational meltdown through accruement of deleterious load. This thesis explores these LD-linked issues in two species of aspens and poplars (Populus) and presents solutions to these problems with the aim of establishing small populations as a feasible option in breeding.

Rami-Petteri Apuli received his graduate education at Oulu university, Oulu, Finland, and acquired a M.Sc. in Biology with specialization in Biosciences and major in Genetics and Physiology.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

Online publication of thesis summary: http://pub.epsilon.slu.se/

ISSN 1652-6880
ISBN (print version) 978-91-7760-797-7
ISBN (electronic version) 978-91-7760-798-4