UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 620

Adaptive Evolution and Demographic History of Norway Spruce (*Picea abies*)

THOMAS KÄLLMAN

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2009

ISSN 1651-6214
ISBN 978-91-554-7457-7
urn:nbn:se:uu:diva-99262

Dissertation presented at Uppsala University to be publicly examined in Zootissalen, EBC, Villavägen 9, Uppsala, Friday, April 24, 2009 at 10:00 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

Källman, T. 2009. Adaptive Evolution and Demographic History of Norway Spruce (Picea abies). Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 620. 36 pp. Uppsala. ISBN 978-91-554-7457-7.

One of the major challenges in evolutionary biology is to determine the genetic basis of adaptive variation. In Norway spruce (*Picea abies*) the timing of bud set shows a very strong latitudinal cline despite a very low genetic differentiation between populations.

The timing of bud set in Norway spruce is under strong genetic control and triggered by changes in photoperiod, but no genes controlling this response have so far been described. In this thesis we used a combination of functional studies to identify candidate genes, and analyses of DNA sequence polymorphism to infer demographic history and test candidate genes for signs of selection.

By monitoring gene expression during bud set in two populations with divergent bud set response, more than 3300 genes were differentially expressed. The response was very similar in the two populations and no significant expression differences were found between populations. The majority of genes showed a gradual change in expression over time, but a small group of around 300 genes showed a strong increase in gene expression already after a single long night. Among these we found genes that are similar to photoperiod pathway genes in Arabidopsis and genes that have been assigned to stress- and cold-response in flowering plants. To further investigate the role of photoperiodic related genes in bud set response detailed expression pattern of four genes, *PaFT1*, *PaFT2*, *PaFT3* and *PaFT4*, homologous to Arabidopsis *FLOWERING LOCUS T* (*FT*) and *TERMINAL FLOWER 1* (*TFL1*) were monitored under more elaborate experimental conditions. *PaFT4* showed an expression pattern correlating with bud set under several different light and photoperiod treatments, indicating that it has a role in the induction of bud set in response to changes in photoperiod.

To investigate the role of selection, demography and hybridization in the evolution of spruce species multilocus data sets of DNA sequence data were collected from Norway spruce and three North American spruce species (*Picea breweriana*, *Picea glauca*, *Picea mariana*). In general, few fixed and a large number of shared polymorphic sites were found among the four species. By employing an "Isolation with migration" model to the data, it was clear that most of the shared polymorphisms were retained ancestral variation and evidence of gene flow was only found between Norway spruce and *Picea glauca*. Despite the large number of shared polymorphisms, the actual level of diversity in Norway spruce was lower compared to expectations from a species with continental wide distribution range. The low diversity was coupled with a skewed allele frequency distribution with far more singletons than expected under neutrality, suggesting that fluctuating population sizes have had a large impact on the diversity observed today. The observed pattern at most genes seems to be consistent with an ancient and sever bottleneck.

To examine the role of selection at three photoperiodic related genes, an approach where the effect on genetic variation due to demographic events was taken into account. One of the genes, *PaPRR3* deviated significantly from expectations based on the inferred demographic scenarios and hence might have been affected by selection.

In summary, this thesis shows that, by using several approaches, we might be able to identify genes involved in local adaptation even in non-model systems, like conifers.

Thomas Källman, Evolutionary Functional Genomics, Department of Evolution, Genomics and Systematics, Uppsala University, SE-75236 Uppsala, Sweden

ISSN 1651-6214
ISBN 978-91-554-7457-7
urn:nbn:se:uu:diva-99262 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-99262)

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Källman T., Ralph S.G., Gyllenstrand N., Lascoux M., Clapham D., White R., Bohlmann J. and Lagercrantz U. (2009) Changes in gene expression during bud set in Norway spruce. Manuscript

II Gyllenstrand N., Clapham D., Källman T. and Lagercrantz U. (2007) A Picea abies FT Homolog Is Implicated in Control of Growth Rhythm in Conifers. *Plant Physiology* 144: 248-257

III Chen J.*, Källman T.*, Gyllenstrand N. and Lascoux M. (2009) New insights on the speciation history and nucleotide diversity of three boreal spruce species and a Tertiary relict. Submitted Manuscript

IV Heuertz M.*, De Paoli E.*, Källman T.*, Larsson H., Jurman I., Morgante M., Lascoux M. and Gyllenstrand N. (2006) Multilocus Patterns of Nucleotide Diversity, Linkage Disequilibrium and Demographic History of Norway Spruce [Picea abies (L.) Karst]. *Genetics* 174:2095-105

V Källman T., De Mita S., Larsson H., Heuertz M., Lagercrantz U. Gyllenstrand N. and Lascoux M. (2009) Identification and evolution of pseudo response regulators in the conifer Norway spruce. Manuscript

* Joint first authors.

Contents

1. Introduction

At first glance, conifers do not appear to be an ideal system for unraveling the genetic basis of adaptive variation. Their genomes are huge and poorly known, the genomic resources a far cry from those at hand in model organisms such as *Arabidopsis thaliana* or *Drosophila*, and their generation time hopelessly long. Yet the situation may not be as bleak as it seems. First, conifers often have huge distribution ranges and large effective population sizes ([1], Paper III). It implies that natural selection, if present, will be very efficient and its footprint in the genome more easily detectable. This is, for example, in contrast to the situation in humans that have more limited genetic variation and a much smaller effective population size [2]. Second, especially in species of major economic importance such as Norway spruce one benefits from decades of provenance tests, progeny testing and growth chamber experiments. In other words, both the phenotypic and quantitative genetic variation is very well characterized. Third, surveys of genetic variation at previously popular markers, be it allozymes, microsatellites or chloroplast PCR-RFLP, were often carried out on a grand scale, involving tens of populations and thousands of individuals (e.g., [3]). These massive efforts, combined with equally impressive efforts by paleoecologists, have endowed us with a rather rich and detailed picture of the history of the species over the last 20,000 years or so. Last but not least, the combination of these different sources of information has left us with apparent paradoxes begging for answers. How can the lack of genetic differentiation among populations observed at neutral markers be reconciled with the steep clines in budset or other adaptive traits? If trees indeed arrived in Scandinavia less than 10,000 years ago as suggested by the pollen fossil records, i.e., less than 200 generations years ago, how can they be so strongly locally adapted? Was selection on adaptive traits really that strong? How would this strong selection on adaptive traits translate at the loci underlying the traits?

The present thesis is a modest attempt to start addressing some of these questions in Norway spruce (*Picea abies*). The trait we have chosen to focus upon is timing of budset, a character of obvious adaptive value and on which a wealth of information is already available (see for example [4, 5]). There are two major approaches to understand the genetic basis of adaptive traits [6]. In the top-bottom approach, candidate genes are first identified through Quantitative Trait Loci (QTL) studies or association mapping and their polymorphisms are analyzed for signatures of adaptation. A striking example of this approach is the identification of the nucleotide polymorphism responsible for plate-morph variation in sticklebacks [7]. The bottom-up approach starts

by identifying signatures of adaptation across the genome through population genetics studies, and then proceeds by analyzing the identified genes functionally in order to confirm their association to the phenotypic variation of interest. The two approaches can converge on overlapping sets of genes as was beautifully demonstrated by Wright *et al.* [8] in maize. This latter study was also noteworthy by the elegant method it developed to account for the demographic history of maize. Accounting for demographics when testing for the presence of natural selection on a gene is crucial as both selection and demographic events can leave the same signature. In the present thesis we first carried out gene expression studies to identify good candidate genes for budset. Candidate genes were also chosen based on information available on orthologous genes in the model plant *A. thaliana*. Since budset has been known for a long time to be under the control of photoperiod, candidate genes were chosen in the photoperiodic pathway, one of the major pathways contributing to the control of flowering time in *A. thaliana* [9]. Second, we used both candidate and randomly chosen genes to try and identify genes undergoing adaptive evolution. Most of the articles included in the thesis are dealing with Norway spruce. However, comparative genomics offer promising new avenues for the understanding of adaptation, so we also started analyzing DNA polymorphism in other species, with similar or divergent histories.

1.1 The conifer *Picea abies*

1.1.1 Phylogeny of conifers

Spruce species (*Picea*) are phylogenetically most closely related to *Pinus*, both of which are included in the order Pinales (formerly known as Coniferales), today the largest group of gymnosperm species. In total around 630 different species of conifers have been described and the family *Pinaceae* is the largest with more than 200 members, including both *Picea* and *Pinus*. Within *Picea* the number of present day species are between 30 and 56 depending on the accepted classification [10, 11]. Spruce species are mainly found in temperate regions of the northern hemisphere and the species with more southern distribution ranges occur mainly at high altitudes. The majority of species are found in Asia, many of them around the Tibetan plateau and only three species are found in Europe, namely the widespread *P. abies* in the Alps and across all Northern Europe, *P. obovata* in Eastern Russia and extending deep into Siberia and the relict *P. omorika* in small pockets in the Balkans. The phylogenetic relationships between spruce species have not been fully resolved and depending on the markers used to reconstruct phylogeny different species trees have been obtained. Early attempts to reconstruct the relationship used morphological characters, but did not manage to resolve the relationship within the spruce species [11]. Molecular markers have led to a division into 3 or 4 groups, which in turn

does not reflect the groups obtained by morphological characters [12]. At present, we do not have a completely reliable phylogeny of spruce species.

1.1.2 The Genomes of conifers

Most species of the extent gymnosperms have large genome sizes, conifers being no exception. In *Picea* the size varies between 15 and 20 Gb, that is around 100 times the genome size of the model plant *Arabidopsis thaliana* [13]. Contrary to what has been observed in angiosperms the increase in genome size is not due to polyploidization, but rather due to an increase of repetitive DNA, mainly in the form of transposable elements, like retrotransposones. Reassociation studies indicate that around 75% of the genomes of *Picea* and *Pinus* are moderately to highly repetitive [14]. Only a few studies have focused on the non-genic sequence regions of conifers, but recent sequencing of genomic DNA from *Picea* and *Pinus* species have shown that the retroelements were mainly active in a common ancestor to the *Pinaceae* group, even though some of the genome size variability within *Pinaceae* could be explained by genera and possibly species specific retroelements [15, 16].

Given the large genome size of conifer species there is still no available genome sequence from any conifer species, something that might change in the near future (http://pinegenomeinitiative.org). Still, even in the absence of complete genome data in conifers, sequencing efforts mainly from expressed parts of the genome have today led to more than 1 million expressed sequence tags (EST) available in public databases. Analysis of EST sequence data from both *Pinus* and *Picea* have revealed close to 50 000 putative unique transcripts and even among the full length sequences obtained, more than 10% did not have obvious homologs in angiosperm species [17, 18]. These EST sequences have not only been useful as a sequence resource, but have also been utilized to create microarray platforms, to study changes in gene expression on a genome wide scale [19, 20], (Paper I). Besides a fairly good knowledge about the expressed part of the conifer genomes there are also a number of comparative mapping studies of conifers. Both karyotypic analyzes and more detailed studies using either Random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLP), microsatellites and more recently sequence or single nucleotide polymorphisms (SNP) suggest that synteny is high and large scale rearrangements uncommon within conifers [21, 22, 23].

1.1.3 Phylogeography of Norway spruce

Since large parts of Northern Europe have gone through glacial cycles during the last hundreds of thousands of years the distribution of species has strongly varied over time (see for example [24]). The last glacial maximum in Europe was around 18000 years ago, when more or less all of Scandinavia and large parts of continental Europe were covered by ice [25]. This pushed many present day species down to more southern latitudes. However, both genetic data and fossil data suggest that many cold-tolerant species likely had a

fairly large distribution range even during glacial periods [26]. Nonetheless, the present day populations of Norway spruce can be divided into three main groups, the Baltico-Nordic, Alpine and finally the Hercyno-Carpathian, corresponding to the inferred refugia, from which Europe was recolonized after the last glaciation [3, 27, 28]. Within groups there is generally limited population genetic structure, but significant, albeit low F_{ST} values have been reported for both nuclear microsatellites and mtDNA in the Baltico-Nordic and Alpine domains [29].

1.1.4 The control of bud set and bud burst in Norway spruce

Plants growing in temperate regions of the world need to adjust their yearly growth rhythm to seasonal conditions. In Norway spruce, as well as in other perennial plants, this means that they have an active growth phase during the summer and a dormant phase during the winter [30]. Starting in spring, an increase in temperature leads to bud burst, which initiates shoot extension. In late summer or early fall short days will induce growth cessation and bud set and the trees enter endodormancy. After an extended period of temperatures around 0°C the plants reach the ectodormant or quiescent state, where they can withstand very low temperatures and resume growth if temperature is permissive (Figure 1.1). These transitions between active growth and dormancy

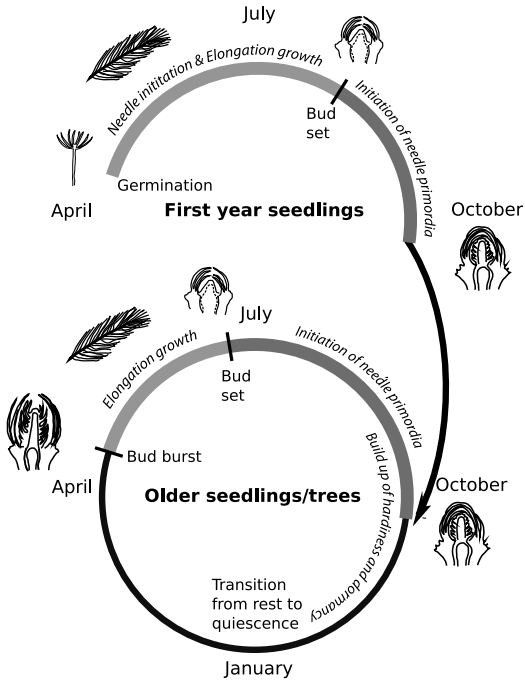

Figure 1.1: A summary cartoon of the annual growth cycle of Norway spruce.

are crucial for plant survival and fitness. For example, if a plant stops grow-

ing at a premature stage it will miss a substantial part of the growing season, whereas a late response might lead to frost damage. The timing of transition between active growth and dormancy is mainly controlled by photoperiod in the autumn and by temperature in the spring [31, 32]. The shift between these two states is under strong genetic control [4]. Physiological experiments, using populations from different latitudes have shown that the critical night length for bud set response differs between populations from different latitudes [31]. The very northern populations stop growing at night lengths of two to three hours, whereas the southern populations (Northern Italy and Romania) need close to 10 hours of night to initiate growth cessation and bud set (Figure 1.2).

Based on the trees response to different light treatments Clapham *et al.* [5] have suggested that phytochromes are likely to play an important role in bud set response to photoperiod, but there is still a dearth of studies investigating the molecular and genetic basis of bud set response in conifers. In the model plant *Populus*, two genes previously known as flowering time genes in Arabidopsis, *FLOWERING LOCUS T* (*FT*) and *CONSTANS* (*CO*) were shown to be involved in bud set and annual growth rhythm [33]. This means that at least between angiosperm trees and annual plants, the genetic pathways involved in light response might show substantial conservation.

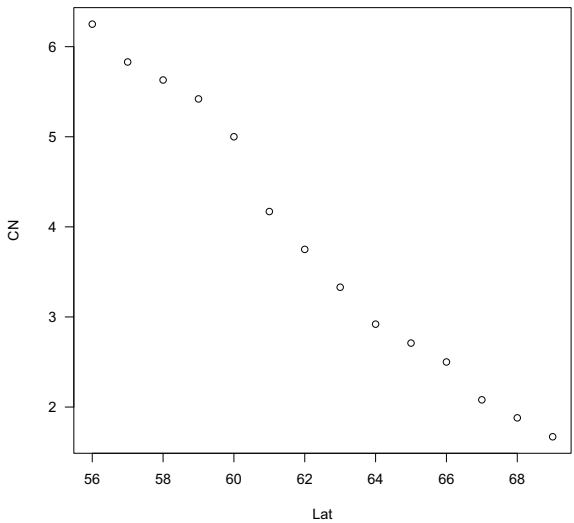

Figure 1.2: Critical night length requirements (CN) for bud set in populations of *P. abies* from different latitudes. Data from [34]

1.2 Identifying genes involved in local adaptation

A natural first step in identifying genes involved in local adaptation is to dissect phenotypic traits involved in local adaption into their genetic components [35]. The majority of traits involved in local adaptation are likely controlled by several loci and environmental interactions. Two main approaches are currently routinely used for dissection of quantitative traits. Quantitative trait loci (QTL) mapping has been a successful approach in species where it is possible to create crosses between divergent individuals and then track the co-segration of genetic variants and phenotypes. More recently, however, association mapping has emerged as a fundamental tool for dissecting complex trait variation. Association mapping is, just like QTL mapping, based on linkage disequilibrium between markers in the genome, but utilizes the whole recombinational history of the studied populations. This leads to the in general much higher resolution in association mapping compared to QTL mapping. The success rate of association mapping depends on knowledge about the fundamental population genetic properties, such as level of diversity, patterns of linkage disequilibrium and the demographic history of the species [35, 36].

In conifers we still have limited knowledge about these parameters. Nonetheless, successful associations have been reported [37], but the results have not yet been validated with functional data. Furthermore, since no genome data is available for any conifer, and genome wide patterns of variation are not available, candidate gene approaches are likely to be the main tool in association mapping in the foreseeable future. Identification of candidate genes has in conifer studies often been restricted to genes homologous to interesting genes from Arabidopsis (see for example [38], Paper IV) Association mapping of candidate regions has been a successful method in several organisms, where one has often started with either detailed knowledge about genetic pathways controlling a trait and/or QTL regions from which genes were chosen. Since the evolutionary distance between angiosperms and gymnosperms is large, relying on sequence homology alone to identify candidate genes might be problematic. In addition, such a strategy will miss any novel aspects where either the process of interest or the genes do not exist in any model organisms.

An alternative approach to identifying genes involved in local adaptation is to use population genetic data sets to identify genes displaying variation that deviates from either empirical genome wide distributions or from simulated data capturing the expected variability in the genome [39, 40]. However, while genes showing a significant deviation might be under selection it does not imply that the gene is involved in local adaptation. Therefore, this approach need to be coupled with functional evaluation to be useful in the identification of adaptive evolution.

1.3 Making sense of multilocus sequence data

The main theoretical framework used for analyzing population genetic data today is known as the coalescent and was formalized mathematically by Kingman [41, 42]. In short, it describes the process whereby alleles in a finite population meet, or coalesce, as we go backwards in time. In a small population, the coalescence rate, which is inversely proportional to the effective population size, will be high and the most common recent ancestor to all observed alleles will be reached quickly, whereas in a large population the coalescence rate will be small and the most common recent ancestor will tend to be found further back in time. The standard coalescent or standard neutral model, has a number of fairly demanding assumptions such as random mating and an unstructured population of constant size over time. Furthermore, alleles are assumed to be neutrally evolving. Still, predictions made with the model are fairly robust and have proven to be useful even in systems which clearly do not conform to the assumptions of the model. For example, under the standard neutral model both the number of segregating sites (S), divided by a constant depending only on the sample size (a_1), and the average pairwise nucleotide difference (π) are unbiased estimates of the population mutation rate, (θ). Tajima used these properties to construct his test of neutrality, Tajima´s D (Equation 1.1) [43]. Hence if a sample has a Tajima´s D value significantly different from zero, one can investigate which assumption is violated and by this come a step closer to understanding how diversity is shaped in the sample.

$$D = \frac{\pi - S/a_1}{\sqrt{\hat{var}[\pi - S/a_1]}} \tag{1.1}$$

Tajima´s D was historically the first of a family of tests using the allele frequency spectrum to test for deviations from the standard neutral model. Some of these tests, like Fay and Wu´s H, also utilize the folded frequency spectrum, that is, frequency of derived alleles in relation to the ancestral state. The latter can be determined with sequence from a closely related species [44, 45] (Equation 1.2).

$$H = \pi - \theta_H \tag{1.2}$$

where θ_H is the homozygosity of derived variants. More recent theoretical work and extensive simulation work have highlighted the fact that the power of the tests can be increased by combining different test statistics that are sensitive to different aspects of frequency deviations [46]. There have been several extensions made to the standard coalescent and we can now easily simulate complex demographic scenarios that include both recombination, population structure and growth [47].

There has recently been a rapid progress in computational methods to analyze multilocus DNA polymorphism data [48, 49, 50, 40]. Many of the commonly used programs make use of the ease with which population genetic data can be simulated. For example, several programs use Approximate Bayesian

Computation (ABC), where parameters are estimated using an rejection algorithm. In brief, model parameters are drawn at random from a prior range and fed into software for coalescent simulations. Summary statistics from the simulation is then compared to the observed summary statistics and if the discrepancy between the simulated and observed values are small enough the parameter values used in the simulation are recorded as a sample from the posterior distribution. The procedure is then repeated until a large number of accepted values has been obtained. ABC approaches have now been used in several plant specie to analyze multilocus data. For example, an ABC approach was used to compare and evaluate different demographic models in Arabidopsis and *Populus tremula* [51, 52]. In both of these applications simulations within the ABC framework were also used to evaluate the relative support for different demographic models.

1.4 Research aims

The thesis had two main aims. First to to identify genes involved in growth cessation and bud set in Norway spruce, and second, to investigate the role of selection, hybridization and demographics in shaping present day sequence diversity in Norway spruce.

Specific goals

I Characterize global gene expression pattern in Norway spruce during growth cessation and identify candidate genes for bud set response (Paper I).

II Investigate the expression of *PaFT4* during bud set and bud burst in populations of Norway spruce with divergent bud set response (Paper II).

III Clarify the origin of extensive shared polymorphisms observed between spruce species from Europe and North America (Paper III).

IV Characterize patterns of nucleotide diversity and linkage disequilibrium in Norway spruce (Paper IV).

V Evaluate the role of selection in the evolution of putative circadian clock related genes in Norway spruce (Paper V).

2. Results and Discussion

2.1 Identification of candidate genes for bud set control in Norway spruce

In paper I and paper II we used expression data to identify genes involved in, and possibly controlling, bud set and bud burst in Norway spruce. We first monitored global changes in gene expression over the complete bud set process using microarrays. Based on these results we did a more detailed study of gene expression of a few selected genes under a much wider range of photoperiodic conditions and under different light treatments. In both studies two populations of Norway spruce, one from northern Sweden and one from Romania, that had been shown in previous studies [5, 31] to have a divergent response to shifts in photoperiod, were cultivated for 10-14 weeks under constant light condition and then transferred to different photoperiod treatments. The two studies led to the identification of a large group of genes that can be used as candidate genes in upcoming association studies.

2.1.1 Global patterns of gene expression during bud set - Paper I

In paper I we used global expression data to describe the complete bud set process in Norway spruce. Samples were collected at 6 time points, one day before the transfer to short days and 1, 4, 7, 14, and 28 days after the shift to short days. Using cDNA microarrays based on sequences from *Picea sitchensis* and *Picea glauca* we investigated patterns of gene expression in more than 21 000 genes. In total we identified 3377 significant genes over the time series, but no significant difference was observed between the two divergent populations. However, by clustering genes with similar expression patterns over time one could note that the average response in level of gene expression was often slightly stronger and faster in the northern population consistent with the faster bud set response observed for this population. The majority of significant genes showed a gradual change over time, but two groups of genes, in total 322 genes, showed a strong response already after the first long night. Among those genes we found *PaFT4*, a gene similar to the *FLOWERING LOCUS T* (*FT*) and *TERMINAL FLOWER 1* (*TFL1*) from angiosperm that have been implemented not only on flowering time regulation, but also in bud set response in (Figure 2.1) *Populus* [33]. Together with *CONSTANS* (*CO*) genes, *FT* has been suggested to be a key components of the photoperiodic pathway. Due to this key role in photoperiodic response we looked in more detail at the expression pattern of genes homologous to Arabidopsis photope-

riod related genes. Two *CO*-like genes were present on the microarray, but none of them showed altered expression pattern over time. Three genes similar to *HAP* genes, which are known to interact directly with *CO* in Arabidopsis were however differentially expressed, and a homolog to the MADS-box gene *SUPPRESSOR OF CONSTANS 1* (*SOC1*) that acts downstream of *FT* was up regulated in response to short days (Figure 2.1). It is hence possible that part of the genes involved in photoperiodic response in bud set in *Populus* is conserved in conifers.

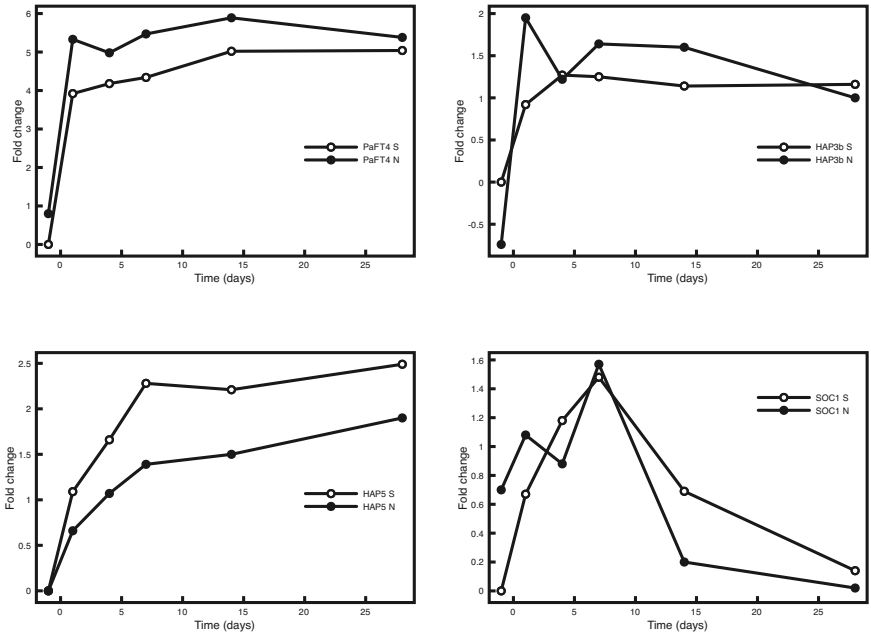

Figure 2.1: Expression pattern of a selected group of photoperiod related in short day induced bud set in Norway spruce.

Another group of genes showing altered expression patterns when shifted to short day conditions includes genes typically assigned to dormancy and cold acclimatization, like dehydrins and *DORMANCY ASSOCIATED PROTEIN* (*DRM*). Homologs to the latter have been used to determine if *Pisum* plants are dormant, and are also up-regulated in dormant tissue in Arabidopsis and *Populus* [53, 54, 55]. The eight dehydrins that were up-regulated in our experiment were in general down-regulated during bud burst in Norway spruce [56]. This suggests that part of the genes involved in light induced dormancy in autumn could be active throughout the dormant period and are then down regulated during spring in response to increased temperature in agreement with a common genetic pathway for temperature and light response.

2.1.2 The role of PEBP-like genes in growth rhythm - Paper II

In this study we focused on the expression pattern of four different genes, *PaFT1*, *PaFT2*, *PaFT3* and *PaFT4*, related to the PEBP gene family in Arabidopsis. This was done partly because *PaFT4* was found to be differentially expressed over time in the microarray experiment (Paper I), but also because it has been shown that homologs to *FT* are key components in the photoperiodic control of bud set in *Populus* [33]. In order to evaluate the role of FT-like genes in bud set, we looked at gene expression under four different photoperiodic conditions as well as under a number of additional light treatments. The four photoperiodic treatments were short day conditions (8h light, 16h dark), long day conditions (19h light, 5h dark), short days with a late night break and short days with an early night break, and they were chosen to maximize differences in bud set response between populations.

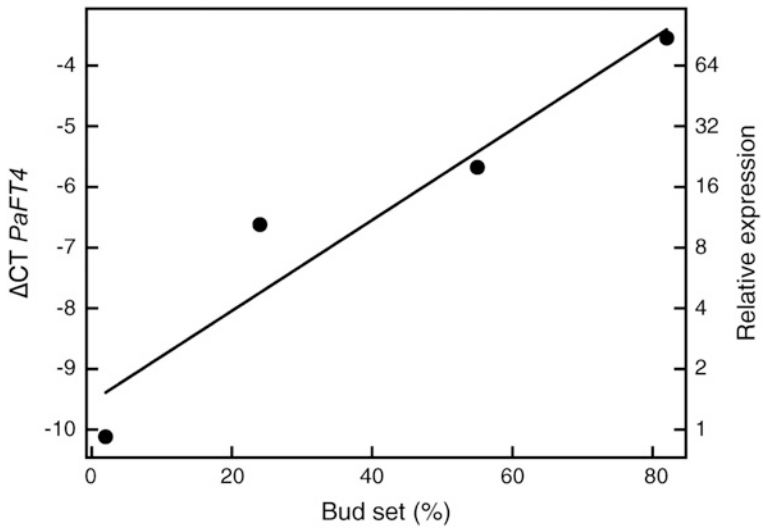

Figure 2.2: Correlation between bud set response and expression level of *PaFT4* under the four different photoperiodic treatments for Romanian population of Norway spruce

As expected, since the northern population requires only a few hours of dark period, close to 100% bud set frequency was observed in all treatments in this population. This was coupled with a very strong increase of expression of the gene *PaFT4*, whereas none of the other PEBP-like genes showed an altered expression pattern over time. In the southern population bud set frequency varied with the treatment and only short day conditions led to a high frequency of bud set, supporting the fact that long night is needed to obtain bud set in southern populations of Norway spruce. Furthermore, the expression pattern of *PaFT4* showed a strong correlation with bud set response under the different treatments in the southern population (Figure 2.2).

Interestingly, the expression pattern of *PaFT4* did not only correlate with bud set, but also showed a decrease in expression pattern during bud burst,

which is mainly under the control of temperature. Hence, both light and temperature seems to affect the expression pattern of *PaFT4*, suggesting that both light and temperature cues from the environment are integrated in the response of the gene *PaFT4*. In summary, there seems to be substantial conservation in genes involved in growth cessation in gymnosperms and angiosperms. However, there might also be novel genes involved in the process that lack homology to known genes from Arabidopsis, begging for further functional analysis of these putative elements, before they can be considered as candidate genes for timing of bud set. In conclusion, a group of genes including *PaFT4*, and dehydrins show strong up-regulation during bud set and down-regulation during bud burst. These genes would be interesting to integrate as candidate genes in both expression, sequence and association mapping studies.

2.2 Multilocus pattern of divergence and diversity in spruce species

In paper III to V we investigated patterns of nucleotide diversity in four different spruce species. In Norway spruce we sequenced 28 different genes covering more than 20kb of aligned sequence from close to 50 individuals. Ten of the fragment were also sequenced in *P. breweriana*, *P. glauca* and *P. mariana*. *P. glauca* and *P. mariana* are, like Norway spruce, boreal species with very large distribution ranges and are common in large parts of Canada and Alaska. *P. breweriana*, on the other hand has a very restricted distribution range in Northern California (Figure 2.3).

Figure 2.3: Present day natural distribution range for *P. abies*, *P. breweriana*, *P. glauca* and *P. mariana*. Note that the distribution range of *P. glauca* and *P. mariana* is largely overlapping and is represented by single shaded area.

Despite these large differences in present day distribution, all four species have very similar levels of diversity at the isoenzyme level [57]. However, even though the variation was not an order of magnitude lower compared to the boreal species, as one would expect from current day distribution, it is clear that the nucleotide diversity of the boreal species was much higher

Table 2.1: *Observed multilocus estimates of nucleotide diversity from different plant species. References to the studies can be found below the table*

Species	π_{All}	π_s	Number of Loci	Sample size
Arabidopsis thaliana	7.0	9.0	374	20
Arabidopsis lyrata	14.1	22.5	77	95
Oryza sativa	2.3	3.2	111	72
Oryza rufipogon	3.6	5.2	111	21
Populus tremula	4.2	8.4	77	12, 19
Zea mays parviglumis	9.5	nr [a]	774	12
Zea mays mays	6.4	nr [a]	774	11
Pinus taeda	4.0	6.4	19	32
Pinus sylvestris	4.0	6.5	16	40
Picea abies	2.5	4.5	28	44
Picea glauca	5.3	9.1	10	22
Picea mariana	3.2	4.6	10	22
Picea breweriana	0.9	2.0	10	22

a) nr: not reported. π_{All}: Total nucleotide diversity, π_s: Silent nucleotide diversity

Arabidopsis thaliana [58], *Arabidopsis lyrata* [59], *Oryza sativa* [60], *Oryza rufipogon* [60], *Populus tremula* [52], *Pinus taeda* [1], *Picea abies* Paper IV and V, *Picea glauca* Paper III, *Picea mariana* Paper III, *Picea breweriana* Paper III

than in *P. breweriana*. Nonetheless, compared to several other angiosperm species the diversity in the boreal spruce species were in the lower range (Table 2.1). Since multilocus patterns of Tajima´s D (Figure 2.4) in all boreal species were skewed towards negative values, a possible explanation to the low diversity estimates could be deviations from the standard neutral model. Despite these lower than expected values of diversity, shared polymorphisms were very common and fixed differences very rare among the three boreal spruce species.

With this data at hand we addressed three main questions, which will be discussed in the coming three sections.

1. What is the origin of observed shared polymorphic sites, are they due to recent migration events or ancestral variation?
2. Could the apparently low nucleotide diversity in the widespread Norway spruce be due to the demographic history of the species?
3. Is diversity observed in photoperiodic related genes compatible with the inferred demographic scenarios?

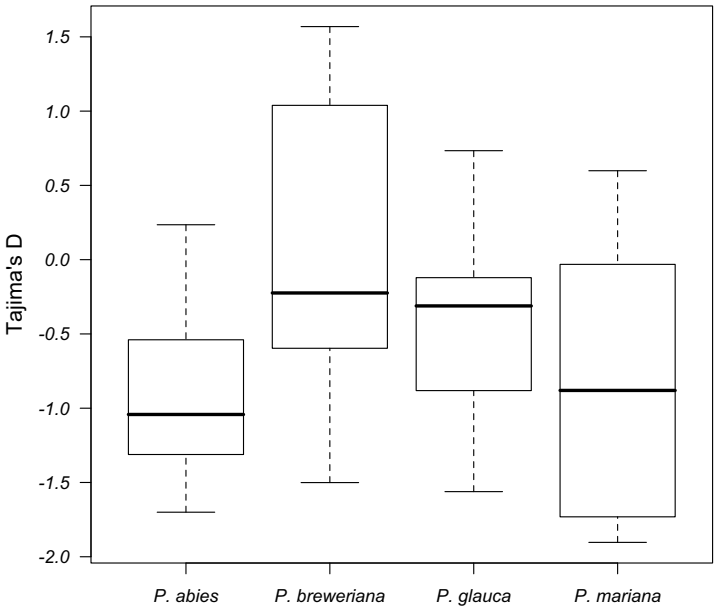

Figure 2.4: Observed Tajima´s D values for the four studied species *P. abies*, *P. breweriana*, *P. glauca* and *P. mariana*.

2.2.1 Patterns of diversity and speciation in four spruce species - Paper III

In order to evaluate the origin of the large number of shared polymorphic sites observed between the three boreal species, we utilized an Isolation-Migration (IM) model that allows for recombination [61]. Briefly, the model estimates six parameters in a pair of diverging populations or species: the divergence time, T, the current effective population sizes of the two species, N_1 and N_2, as well as the effective population size of the ancestral population from which they derived, N_A, and directional gene flow occurring after the two species diverged, M_1 and M_2. Once estimates of parameters were obtained, the fit of the resulting model to the data was tested using a goodness-of-fit framework. Simulated data were obtained by randomly drawing values of the parameters from their inferred posterior distributions and using those in a coalescent simulator [47, 61]. A set of summary statistics was then calculated from the simulated data and compared to the observed values. The fit for nucleotide diversity , F_{ST}, numbers of private, shared and fixed alleles was good while the fit between observed and simulated values for Tajima´s D was poorer. This came as no surprise since the observed mean values of Tajima´s D in all three boreal species was negative and the standard implementation of IM models

assumes a standard neutral model within species. Since the average Tajima´s D was negative in the three boreal species we tested an IM model where we introduced population growth for the boreal species but kept a standard neutral model for *P. breweriana*. This significantly improved the goodness of fit of the model to the data.

Some very interesting results emerged from these data. First, boreal species have fairly large effective population sizes, though not as large as the first estimate of N_e that was proposed for a conifer species based on nucleotide diversity data [1]. Second, nucleotide diversity in *P. breweriana* was an order of magnitude lower than in the three boreal species and the effective population size was also much smaller. This is in stark contrast to the rather similar levels of heterozygosity observed at allozyme loci in the different species. Discrepancy between levels of nucleotide and allozyme diversity have been observed before. Because the species investigated here had very different demographic histories, the discrepancy might well be due to selection as suggested by Pyhäjärvi *et al.* [62]. Third, we found evidence of gene flow between *P. glauca* and Norway spruce. Since the two species cannot be crossed easily today and are found on different continents, the observed gene flow could possibly reflect the sympatry of the two species during a warm period when spruce species were common in Greenland [63].

Overall, the results from this study highlights the presence of shared ancestral polymorphisms in spruce species, despite a divergence time of around 10-20 millions years. In retrospect, this is not surprising since the effective population size of conifers in general is large and they have long generation times, reducing the 10-20 millions years to a few coalescent time units. In other words, drift in species with very large effective population sizes works slowly. Our study, as other before ([64] and references therein), hence highlight the need to take this into account when investigating the phylogenetic relationships between species.

2.2.2 The role of demography in the evolution of Norway spruce - Paper IV

In paper IV we used a multilocus data set of sequence data from individuals sampled from the natural distribution range of Norway spruce (Figure 2.5) to look at patterns of diversity and at the role of demography in shaping this variability. We used sequence data from a total of 22 genes in close to 50 gametes representing seven natural populations. 11 of the genes were originally identified as genes similar to Arabidopsis photoperiod related genes and the remaining ones were randomly chosen from a Norway spruce EST library. All of them seemed to evolve neutrally. On average, a polymorphic site was found every fifty basepairs and close to 50% of the polymorphisms were singletons. Approximately half of the obtained sequence data was coding sequence and non-synonymous diversity was around a third of the observed average silent diversity. Both levels of diversity and patterns of linkage disequilibrium (LD) were consistent with other studies on conifers, LD decaying rapidly within a

few hundred basepairs [1, 65]. In general, the geographic pattern of diversity could be divided according to the present day populations' proximity to the inferred glacial refugia[3, 27]. This result was obtained both with model based analysis (Structure) and traditional F_{ST} approaches (Figure 2.5).

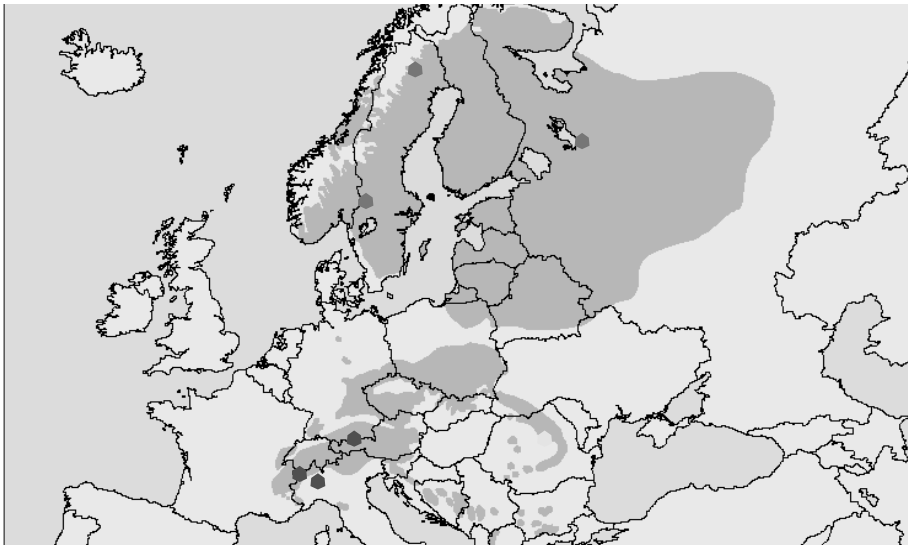

Figure 2.5: Sampled populations of *P. abies* colored according to assigned genetic cluster using the software Structure (Red dots: Baltico-Nordic domain, blue dots: Alpine domain, Yellow dots: Hercyno-Carpathian domain). The blue shading represents the natural distribution range of *Picea abies*.

The lower than expected diversity coupled with negative mean values of Tajima´s D and Fay and Wu´s H (-0.92 and -0.74, respectively) suggest that diversity in Norway spruce might have been curtailed by demographic events resulting in negative values of these two summary statistics. The only demographic scenario that has negative values both for Tajima´s D and Fay and Wu´s H was a fairly old and severe bottleneck. Interestingly, the inferred bottleneck predates the latest glaciation events of Europe (18,000 ya) and even though we did not attempt to time it explicitly, due to the large uncertainty in estimates of mutation rate, the timing coincides with bottleneck events inferred from two other European tree species, *Pinus sylvestris* and *Populus tremula* [65, 52].

2.2.3 Testing for selection at candidate genes - Paper V

Our ongoing functional studies suggested that circadian clock related genes might play a key role in the control of phenological variation (unpublished results). We therefore decided to test for the presence of selection at three genes, named *PaPRR1*, *PaPRR3* and *PaPRR7*, that were identified on the basis of both similarity at sequence level and expression patterns to ARABIDOPSIS PSEUDO RESPONSE REGULATORS (APRR). In both Arabidopsis and

several other angiosperm species the PRR genes have been associated to adaptation to local light conditions [66, 67]. To test for the presence of recent selection at the three genes we proceeded in two steps. First, to account for demographic effects we also resequenced 14 background genes a priori assumed not be involved in local adaptation. Secondly, we used an Approximate Bayesian Computation (ABC) method on these data to evaluate three different demographic scenarios, the standard neutral model (SNM), a population expansion model (PEM) and a bottleneck model (BNM) (Figure 2.6).

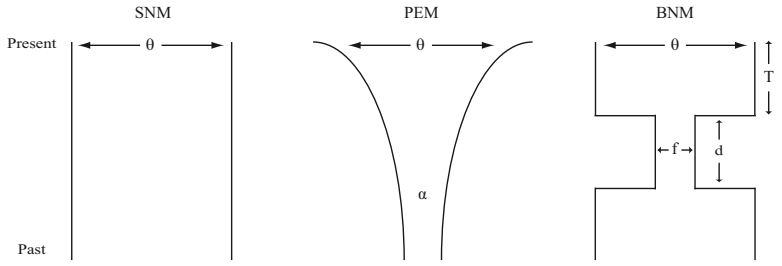

Figure 2.6: Schematic pictures of the three demographic models that were compared by ABC. SNM, Standard Neutral Model, with 2 estimated parameter θ and ρ. PEM, Population expansion model with three estimated parameters ρ, θ and exponential growth rate (α), BNM, Bottleneck model with five associated parameters θ, ρ, times since end of Bottleneck (T), size of population during bottleneck (f), duration of bottleneck (d)

Comparing acceptance rates of the three models suggests that the PEM model outperforms the other two models, even though the difference between the acceptance rate between the BNM and PEM was low. This result differs from the conclusion reached in Paper IV, where, we rejected a simple PEM model and accepted a BNM, based on the negative observed multilocus values of Tajima´s D and Fay and Wu´s H. Given that we here use partly overlapping data we would expect the BNM to better fit the data. The exact cause of this discrepancy is unclear.

We simulated data using a coalescent program and parameter values drawn from the inferred posteriors for all three models. The obtained distribution of summary statistics was compared to the corresponding observed values at the three candidate genes. This gives us the p-values for candidate genes under each of the three models. Under all three models Tajima´s D was significantly higher for *PaPRR3* whereas the summary statistics at the other two genes were closer to the inferred scenarios (Figure 2.7). Interestingly, the variability in *PaPRR3* also show a higher than average geographic structure, which would be consistent with a role in local adaptation.

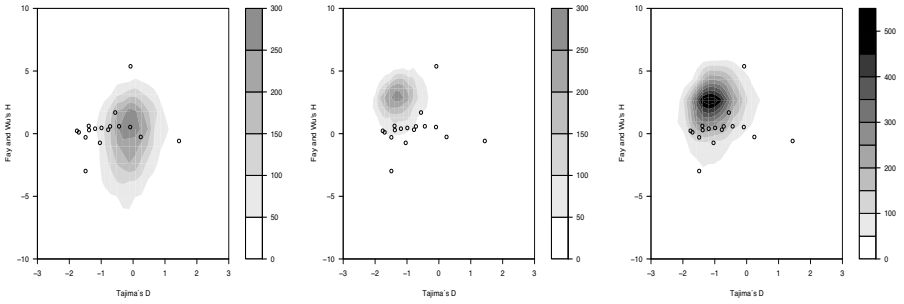

Figure 2.7: Joint distribution of Tajima´s D and Fay and Wu´s H. Shaded area represents the values obtained from simulation under the inferred models and the dots are observed values from the three *PaPRR* genes and the fourteen genes used in demographic inference. From left to right model are SNM, PEM and BNM. Darker shaded regions represents higher density in the posterior distribution.

3. Conclusions

Our studies on gene expression and nucleotide diversity are an encouraging start for future studies on adaptive evolution in Norway spruce and other spruce species. Several physiological studies have earlier highlighted the role of photoperiod in the control of bud set in autumn and of temperature in the control of bud burst during spring. Furthermore, progeny tests showed that this response is largely under genetic control with high estimates of heritability. Despite this, only few genes have been explicitely associated to either bud set or bud burst. In order to identify genes involved in this process we have used a combination of approaches. To identify candidate genes we monitored gene expression on a large set of genes or, on a few, well targeted genes during both bud set and bud burst under controlled environmental conditions.

In order to test for the presence of selection we also collected multilocus sequence data to estimate key population genetic parameters and increase our understanding of the demographic and selection history of Norway spruce. That this general strategy can be fruitful is vindicated by the positive results of the last paper.

There are some possible avenues for future research starting from there. First, future large scale association studies in Norway spruce will hopefully resolve the role of *PaPRR3*, *PaFT4* and other genes in the strong latitudinal cline in bud set commonly observed in Norway spruce. Second, as Paper III showed, information is easily transferred across spruce species. Hence, adaptive genes identified in a given species could be tested in other species. Spruce species are found in very different environments ranging from the vast forests of the Siberian taiga to the scattered high altitude populations of the Himalayas. Data in paper III as well as unpublished results suggest that the different species went through very different demographic histories and they therefore offer a nice opportunity to test the interplay of demography and selection in shaping the current genetic diversity of species. In summary, this thesis shows that, by using several complementary approaches, we might be able to identify genes involved in local adaptation even in non-model systems.

4. Svensk sammanfattning

I stora delar av Europa är gran (*Picea abies*) en av de dominerande trädarterna. Ett flertal studier har påvisat en latitudgradient i knoppsättning och knoppsprickning. Denna gradient beror antagligen på lokala populationer är anpassade till rådande klimatförhållanden. Knoppsättningen styrs framförallt av förändring i dagslängd och induceras under hösten när dagarna blir kortare, medan knoppsprickning på våren framförallt styrs av temperatur. Trots att det finns en stark genetisk kompononent som styr både knoppsättning och knoppsprickning har endast några få gener involverade i denna proccess beskrivits. För att kunna identifiera gener som förklarar lokal anpassning hos populationer, krävs både kunskap om vilka gener som är inblandade i responsen och detaljerad information om ett flertal populationsgenetiska parametrar.

För att identifiera gener som är involverade i trädens respons till förändring i dagslängd undersökte vi förändringar i genuttryck under knoppsättning i två olika populationer av gran, en från norra Sverige och den andra från Rumänien (Artikel I och II). De två populationerna valdes ut, på grund av deras stora skillnad i respons till förändringar i dagslängd. Den norra populationen reagerar med knoppsättning redan vid 2-3 timmar långa nätter medan den södra kräver nattlängder upp mot 10 timmar för att sätta knopp. Genom att följa genuttrycksförändringar från induktion av knoppsättning (långa nätter) till fullständig knoppbildning i dessa två populationer identifierades mer än 3300 gener som uppvisade skillnad i genuttryck över tiden. Trots att populationerna skilde sig åt i knoppsättningsrespons hittades inga statistiskt säkerställda skillnader i genuttryck mellan populationerna. Bland de gener som reagerade kraftigast på förändingen i dagslängd fanns genen (*PaFT4*), som i blommande växter är involverad i initieringen av blomning.

För att i mer detalj undersöka denna typ av gener studerades i Artikel II genutryck av *PaFT4* och tre liknande gener under flera olika dagslängder och ljusbehandlingar. Ingen av de övriga tre generna uppvisar ett uttrycksmönster som samvarierade med knoppsättning, medan uttryckmönstret för *PaFT4* korrelerade med knoppsättning under alla olika dagslängder och ljusbehandlingar. Uttryck av *PaFT4* studerades även i ett knoppsprickningsförsök. Invintrade plantor sattes i ett växthus med gradvis temperaturökning till plantorna påbörjade tillväxt. Uttrycket av *PaFT4* sjönk kraftigt över tiden och var som lägst alldeles efter knoppsprickning. Sammantaget visar resultat från dessa två studier att ett flertal gener involverade i respons till förändringar i dagslängd är bevarade mellan blommande växter och barrträd. Dessutom påverkas flera av generna både av förändringar i dagslängd på hösten och av temperaturförän-

dringar på våren, vilket antyder att det finns en delvis gemensam genetisk bas för temperatur- och ljusrespons i gran.

I de tre följande artiklarna användes populationsgenetiska data från gran samt tre nordamerikanska granarter, slöj-, svart- och vitgran (*Picea breweriana*, *Picea mariana*, *Picea glauca*), för att undersöka effekterna av selektion, förändringar i populationstorlek och hybridisering i evolutionen av dessa arter (Artikel III, IV and V). Mellan de fyra arterna av gran fanns det generellt sett få fixerade genetiska skillnader och en stor del av variationen var gemensam åtminstone i parvisa jämförelser. Denna typ av variation med både fixerade och gemensamma genetiska varianter har visat sig vara användbar för att uppskatta ett flertal populationsgenetiska parametrar, såsom när arterna bildades, om det förekommit genflöde mellan dem sedan artbildningen samt deras effektiva populationsstorlekar. Uppskattning av dessa parametrar från de fyra granarterna visar att merparten av den gemensamma varitionen härstammar från deras gemensamma anfader, en variation som sedan inte fixerats i dagens arter. Ett intressant undantag är dock gran och svartgran som trots att de idag inte går att korsa och dessutom lever på olika kontinenter, uppvisar tecken på genflöde mellan sig. Dock bör man komma ihåg att detta är en uppskattning av genflöde från genetiska data, vilket inte behöver betyda att det finns genflöde mellan dem idag.

Populationsgenetiska analyser visade att granens populationsstorlek antagligen fluktuerat kraftigt över tiden och detta har lämnat ett tydligt mönster i den genetiska variationen vi observerar idag. Den nuvarande granpopulationen visar tydliga tecken på populationstillväxt. Dessa resultat återspeglas i att granen uppvisar mindre genetisk variation än många andra växter trots sin idag mycket stora utbredning och populationsstorlek. Liknande mönster har observerats i flera Europeiska trädarter och verkar inte vara ett resultat av en populationsminskning under den senaste istiden, utan av händelser längre tillbaka i tiden.

I flera växter har gener som styr den interna dygnsrytmen, den så kallade circadiska klockan, visat sig vara viktiga i anpassning till lokala ljusförhållanden. För att undersöka om denna typ av gener också var inblandad i lokal anpassning hos gran studerade vi både genuttryck och genetisk variation i de tre generna *PaPRR1*, *PaPRR3* och *PaPRR7* (Artikel V). De uppvisar alla ett uttryckmönster som liknar det som observerats hos klockgener i blommande växter. För att testa om någon av dessa tre gener påverkats av selektion testade vi den observerade variationen i dessa tre gener mot en modell där vi tog hänsyn till det faktum att granens populationshistoria fluktuerat över tiden. *PaPRR3* uppvisade variation som avvek signifikant från vad som förväntats givet granens demografiska historia och har antagligen varit utsatt för selektion. Resultaten tyder också på lokal anpassing då variationen in denna gen dessutom korrelerar med latitud.

Avslutningsvis ger denna avhandling, genom att studera både genuttryck och genetisk variation, en ökad förståelse för hur knoppsättning i gran regleras på genetisk nivå.

5. Acknowledgements

As obvious from the papers list this is by no means a one mans work, a large number of people have helped me out in the process of writing this thesis. First of all I would like to thank my supervisors Martin Lascoux and Ulf Lagercrantz for scientific guidance and for always having time for a chat even though it in the majority of cases rather have been to discuss things unrelated to science, like latest gadgets from Apple, the poor Swedish food culture (mainly Martin), långfärdsskridskor and orientering (mainly Ulf) and fishing (mainly me).

Second, I want to thank Niclas for helping me get started in the lab and basically being the backbone of the department until you betrayed us and left for SLU. Sharing room with you, Anna-Carin and a five year old "mosbricka" was really a great experience. Anna-Carin thank you for all long and winding discussions on mostly nerdy tech-related stuff and for always being able to help out regarding all kinds of scientific problems. Our room on the third floor, including, spanish homebrew, "nöd-öl", a bunch of ants preserved in ethanol, several old computers and a few stuffed mammals had a nice atmosphere to it that I really miss.

Hanna, even though you seem perfectly happy being at home with your kid, I think you should get back to work, since the place is not the same without you.

Kalle thanks for all interesting political discussions and always approaching topics from a different (strange?) angle.

Many thanks also goes out to Kerstin and Rose-Marie, you are truly people that makes a difference.

This work could not have been done without the people currently or previously associated with the spruce-team at the department - Myriam, Yoshi, Laura, Daniel, David C., Jing, Jun, Xiao-Fei, Li, Michael, Anna, Sofia thank you all.

Many thanks also goes out to Per, Johan, Martin, Sven-Åke, Tanja, Marita, Susanne and others working at the department when I started my PhD and to the present day people at the department, it was back then and still is a fun place to work at.

On the less work related side I also want to thank 3 * Niclas for all the nice fishing trips and this spring I will have full focus on the 10+ pike.

I also want to thank my family back in Jemtland for all support over the years, I would not have managed to get this far without you.

And finally to my favorite girls. Sara och Elsa ni är mitt allt!

Bibliography

[1] Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB: **Nucleotide diversity and linkage disequilibrium in loblolly pine.** *Proc Natl Acad Sci USA* 2004, **101**(42):15255–60.

[2] Przeworski M, Hudson RR, Rienzo AD: **Adjusting the focus on human variation**. *Trends Genet.* 2000, **16**(7):296–302.

[3] Lagercrantz U, Ryman N: **Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation**. *Evolution* 1990, **44**:38–53.

[4] Eriksson G, Ekberg I, Dormling I, Matérn B, von Wettstein D: **Inheritence of bud-set and bud-flushing in *Picea abies* (L.) Karst.** *Theoretical and Applied Genetics* 1978, **52**:2–19.

[5] Clapham D, Dormling I, Ekberg I, Eriksson G: **Latitudinal cline of requirement for far-red light for the photoperiodic control of budset and . . .** . *Physiologia plantarum* 1998, **102**:71–78.

[6] Ross-Ibarra J, Morrell PL, Gaut BS: **Plant domestication, a unique opportunity to identify the genetic basis of adaptation**. *Proc Natl Acad Sci USA* 2007, **104 Suppl 1**:8641–8.

[7] Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM: **Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles**. *Science* 2005, **307**(5717):1928–33.

[8] Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS: **The effects of artificial selection on the maize genome**. *Science* 2005, **308**(5726):1310–4.

[9] Mouradov A, Cremer F, Coupland G: **Control of flowering time: interacting pathways as a basis for diversity**. *Plant Cell* 2002, **14 Suppl**:S111–30.

[10] Farjón A: *Pinaceae: Drawings and Descriptions of the Genera Abies, Cedrus, Pseudolarix, Keteleeria, Nothotsuga, Tsuga, Cathaya, Pseudotsuga, Larix and Picea.* Koeltz ScientiWc Books 1990.

[11] Wright JW: **Species crossability in spruce in relation to distribution and taxonomy**. *Forest Science* 1955, **1**:319–340.

[12] Ran JH, Wei XX, Wang XQ: **Molecular phylogeny and biogeography of Picea (Pinaceae): implications for phylogeographical studies using cytoplasmic haplotypes.** *Mol. Phylogenet. Evol.* 2006, **41**(2):405–19.

[13] Murray BG: **Nuclear DNA Amounts in Gymnosperms**. *Annals of Botany* 1998, **82**:3–15.

[14] Rake AV, Miksche JP, Hall RB, Hansen KM: **DNA reassociation kinetics of four conifers**. *Genome* 1980, **22**:69–79.

[15] Paoli ED: **Diversitá genetica, linkage disequilibrium e compenente ripetitiva del genoma in Abete Rosso (*Picea abies* (L.) Karst.)**. *PhD thesis*, Universitá Degli Studi di Udine 2006.

[16] Morse AM, Peterson DG, Islam-Faridi MN, Smith KE, Magbanua Z, Garcia SA, Kubisiak TL, Amerson HV, Carlson JE, Nelson CD, Davis JM: **Evolution of genome size and complexity in Pinus**. *PLoS ONE* 2009, **4**(2):e4332.

[17] Ralph SG, Chun HJE, Kolosova N, Cooper D, Oddy C, Ritland CE, Kirkpatrick R, Moore R, Barber S, Holt RA, Jones SJM, Marra MA, Douglas CJ, Ritland K, Bohlmann J: **A conifer genomics resource of 200,000 spruce (Picea spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (Picea sitchensis)**. *BMC Genomics* 2008, **9**:484.

[18] Kirst M, Johnson AF, Baucom C, Ulrich E, Hubbard K, Staggs R, Paule C, Retzel E, Whetten R, Sederoff R: **Apparent homology of expressed genes from wood-forming tissues of loblolly pine (Pinus taeda L.) with Arabidopsis thaliana**. *Proc Natl Acad Sci USA* 2003, **100**(12):7383–8.

[19] Ralph SG, Yueh H, Friedmann M, Aeschliman D, Zeznik JA, Nelson CC, Butterfield YSN, Kirkpatrick R, Liu J, Jones SJM, Marra MA, Douglas CJ, Ritland K, Bohlmann J: **Conifer defence against insects: microarray gene expression profiling of Sitka spruce (Picea sitchensis) induced by mechanical wounding or feeding by spruce budworms (Choristoneura occidentalis) or white pine weevils (Pissodes strobi) reveals large-scale changes of the host transcriptome.** *Plant Cell Environ.* 2006, **29**(8):1545–70.

[20] Holliday JA, Ralph SG, White R, Bohlmann J, Aitken SN: **Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (Picea sitchensis)**. *New Phytol* 2008, **178**:103–22.

[21] Pavy N, Pelgas B, Beauseigle S, Blais S, Gagnon F, Gosselin I, Lamothe M, Isabel N, Marie BJB: **Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce**. *BMC Genomics* 2008, **9**:21.

[22] Pelgas B, Beauseigle S, Acheré V, Jeandroz S, Marie BJB, Isabel N: **Comparative genome mapping among Picea glauca, P. mariana x P. rubens and P. abies, and correspondence with other Pinaceae.** *Theor Appl Genet* 2006, **113**(8):1371–93.

[23] Pelgas B, Marie BJB, Beauseigle S, Isabel N: **A composite linkage map from two crosses for the species complex Picea mariana x Picea rubens and analysis of synteny with other Pinaceae**. *Theor Appl Genet* 2005, **111**(8):1466–88.

[24] Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé AE, Martín JP, Rendell S, Vendramin GG: **Glacial refugia: hotspots but not melting pots of genetic diversity.** *Science* 2003, **300**(5625):1563–5.

[25] Clark P, Mix A: **Ice sheets and sea level of the last glacial maximum**. *Quaternary Science Review* 2002, **21**:1–7.

[26] Birks H, Willis K: **Alpines, trees, and refugia in Europe**. *Plant Ecology and Diversity* 2008, **1**:147–160.

[27] Vendramin GG, Anzidei M, Madaghiele A, Sperisen C, Bucci G: **Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (Picea abies K.).** *Genome* 2000, **43**:68–78.

[28] Tollefsrud M, Kissling R, Gugerli F, Johnsen O, Skrøppa T, Cheddadi R, Knaap WVD, Latałowa M, Terhürne-Berson R, Litt T, Geburek T, Brochmann C, Sperisen C: **Genetic consequences of glacial survival and postglacial colonization in Norway spruce: combined analysis of mitochondrial DNA and fossil pollen**. *Mol Ecol* 2008, **17**:4134–4150.

[29] Tollefsrud M, Sønstebø J, Brochmann C, Johnsen O, Skrøppa T, Vendramin G: **Combined analysis of nuclear and mitochondrial markers provide new insight into the genetic structure of North European Picea abies**. *Heredity* 2009 In Press.

[30] Rohde A, Bhalerao RP: **Plant dormancy in the perennial context**. *Trends Plant Sci* 2007, **12**(5):217–23.

[31] Dormling I: **Photoperiodic control of growth and growth cessation in Norway spruce seedlings. International Symposium on Dormancy in Trees**. IUFRO division 2, working party 2.01. 4, Kornik, Poland 1973.

[32] Hannerz M: **Evaluation of temperature models for predicting bud burst in Norway spruce**. *Can J Forest Res* 1999, **29**:9–19.

[33] Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O: **CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees.** *Science* 2006, **312**(5776):1040–3.

[34] Dormling I: **Influence of light intensity and temperature on photoperiodic response of Norway spruce provenances**. In *IUFRO Jt. Meet. Work. Parties Norway spruce provenances Norway spruce breed* 1979.

[35] Lander ES, Schork NJ: **Genetic dissection of complex traits**. *Science* 1994, **265**(5181):2037–48.

[36] Neale DB, Savolainen O: **Association genetics of complex traits in conifers.** *Trends Plant Sci* 2004, **9**(7):325–30.

[37] González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB: **Association genetics in Pinus taeda L. I. Wood property traits**. *Genetics* 2007, **175**:399–409.

[38] González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB: **DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in Pinus taeda L.** *Genetics* 2006, **172**(3):1915–26.

[39] Akey JM, Zhang G, Zhang K, Jin L, Shriver MD: **Interrogating a high-density SNP map for signatures of natural selection.** *Genome Res.* 2002, **12**(12):1805–14.

[40] Thornton KR, Andolfatto P: **Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of Drosophila melanogaster**. *Genetics* 2006, **172**(3):1607–19.

[41] Kingman JFC: **The coalescent.** *Stochastic Processes and their Applications* 1982, **13**:235–248.

[42] Kingman JFC: **On the genealogy of large populations**. *Journal of Applied Probability* 1982, **19A**:27–43.

[43] Tajima F: **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism**. *Genetics* 1989, **123**(3):585–95.

[44] Fu YX: **Statistical properties of segregating sites**. *Theoretical population biology* 1995, **48**(2):172–97.

[45] Fay JC, Wu CI: **Hitchhiking under positive Darwinian selection.** *Genetics* 2000, **155**(3):1405–13.

[46] Zeng K, Fu YX, Shi S, Wu CI: **Statistical tests for detecting positive selection by utilizing high-frequency variants**. *Genetics* 2006, **174**(3):1431–9.

[47] Hudson RR: **Generating samples under a Wright-Fisher neutral model of genetic variation**. *Bioinformatics* 2002, **18**(2):337–8.

[48] Beaumont MA, Zhang W, Balding DJ: **Approximate Bayesian computation in population genetics**. *Genetics* 2002, **162**(4):2025–35.

[49] Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P: **Multilocus patterns of nucleotide variability and the demographic and selection history of Drosophila melanogaster populations.** *Genome Res.* 2005, **15**(6):790–9.

[50] Marjoram P, Tavaré S: **Modern computational approaches for analysing molecular genetic variation data**. *Nat Rev Genet* 2006, **7**(10):759–70.

[51] François O, Blum MGB, Jakobsson M, Rosenberg NA: **Demographic history of european populations of Arabidopsis thaliana**. *PLoS Genet* 2008, **4**(5):e1000075.

[52] Ingvarsson PK: **Multilocus patterns of nucleotide polymorphism and the demographic history of Populus tremula**. *Genetics* 2008, **180**:329–40.

[53] Stafstrom JP, Ripley BD, Devitt ML, Drake B: **Dormancy-associated gene expression in pea axillary buds. Cloning and expression of PsDRM1 and PsDRM2**. *Planta* 1998, **205**(4):547–52.

[54] Tatematsu K, Ward S, Leyser O, Kamiya Y, Nambara E: **Identification of cis-elements that regulate gene expression during initiation of axillary bud outgrowth in Arabidopsis**. *Plant Physiology* 2005, **138**(2):757–66.

[55] Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, Bhalerao RP, Boerjan W, Rohde A: **A molecular timetable for apical bud formation and dormancy induction in poplar**. *Plant Cell* 2007, **19**(8):2370–90.

[56] Yakovlev I, Asante D, Fossdal C, Partanen J, Junttila O, Johnsen O: **Dehydrins expression related to timing of bud burst in Norway spruce**. *Planta* 2008, **228**(3):459–472.

[57] Ledig T, Jacob-Cervantes V, Hodgskiss PD, Equiliuz-Piedra T: **Recent evolution and divergence among populations of a rare Mexican endemic, Chihuahua Spruce, following Holocene climatic warming**. *Evolution* 1997, **51**(6):1815–1827.

[58] Schmid KJ, Ramos-Onsins S, Ringys-Beckstein H, Weisshaar B, Mitchell-Olds T: **A multilocus sequence survey in Arabidopsis thaliana reveals a genome-wide departure from a neutral model of DNA sequence polymorphism.** *Genetics* 2005, **169**(3):1601–15.

[59] Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D, Gaut BS: **Patterns of polymorphism and demographic history in natural populations of Arabidopsis lyrata**. *PLoS ONE* 2008, **3**(6):e2411.

[60] Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR, Bustamante CD, Purugganan MD: **Genome-wide patterns of nucleotide polymorphism in domesticated rice**. *PLoS Genet* 2007, **3**(9):1745–56.

[61] Becquet C, Przeworski M: **A new approach to estimate parameters of speciation models with application to apes**. *Genome Res.* 2007, **17**(10):1505–19.

[62] Pyhäjärvi T: **Roles of demography and natural selection in molecular evolution of trees, focus in *Pinus sylvestris*.** *PhD thesis*, University of Oulu 2008.

[63] de Vernal A, Hillaire-Marcel C: **Natural variability of Greenland climate, vegetation, and ice volume during the past million years**. *Science* 2008, **320**(5883):1622–5.

[64] Edwards SV: **Is a new and general theory of molecular systematics emerging?** *Evolution* 2009, **63**:1–19.

[65] Pyhäjärvi T, García-Gil MR, Knürr T, Mikkonen M, Wachowiak W, Savolainen O: **Demographic history has influenced nucleotide diversity in European Pinus sylvestris populations**. *Genetics* 2007, **177**(3):1713–24.

[66] Michael TP, Salomé PA, Yu HJ, Spencer TR, Sharp EL, McPeek MA, Alonso JM, Ecker JR, McClung CR: **Enhanced fitness conferred by naturally occurring variation in the circadian clock.** *Science* 2003, **302**(5647):1049–53.

[67] Turner A, Beales J, Faure S, Dunford RP, Laurie DA: **The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley.** *Science* 2005, **310**(5750):1031–4.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 620

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-99262

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2009