UNIVERSITÀ
DEGLI STUDI
FIRENZE

DISPAA
DIPARTIMENTO DI SCIENZE DELLE
PRODUZIONI AGROALIMENTARI
E DELL'AMBIENTE


DOTTORATO DI RICERCA IN
Scienze e Technologie Vegetali, Microbiche e Genetiche

CICLO XXVII

COORDINATORE Prof. Aniello Scala

# *Estimation of adaptive genetic variation in Norway spruce (Picea abies (L.) Karst) to climate change*


*PhD candidate: Irina Calic*

December, 2014

## Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

December 31$^{th}$, 2014 Nome Cognome del dottorando

A copy of the thesis will be available at: www.dispaa.unifi.it

## Dichiarazione

Con la presente affermo che questa tesi è frutto del mio lavoro e che, per quanto io ne sia a conoscenza, non contiene materiale precedentemente pubblicato o scritto da un'altra persona né materiale che è stato utilizzato per l'ottenimento di qualunque altro titolo o diploma dell'Università o altro istituto di apprendimento, a eccezione del caso in cui ciò venga riconosciuto nel testo.

31 dicembre 2014 Nome Cognome del dottorando

Una copia della tesi sarà disponibile presso: www.dispaa.unifi.it

# Contents

**Riassunto** ---------- i

**Summary** ---------- ii

**Lavori correlati alla Tesi - Papers related to the Thesis** ---------- iii

**List of Figures** ---------- iv

**List of Tables** ---------- v

**Acknowledgment** ---------- vi

1. **INTRODUCTION** ---------- 1

1.1. About Conifers ---------- 1
1.2 Pinaceae (The Pinus family) ---------- 2
1.3 Norway spruce natural distribution ---------- 2
1.4 Ecology of Norway spruce ---------- 3
1.5 Climate change and its impact on Conifers ---------- 4
1.6 Phenomena of adaptation ---------- 7
1.7 Adaptive genetic variation in Conifers with emphasis on Norway spruce (common garden trials) ---------- 9
1.8 Genotype versus environment in Conifers with emphasis on Norway spruce ---------- 10
1.9 Gene groups associated to environment among Conifers ---------- 12

2. **RESEARCH OBJECTIVES** ---------- 15

3. **MATERIAL AND METHODS** ---------- 16

3.1 Sample collection ---------- 16
3.2 Illumina 384 GoldenGate chip design ---------- 17
3.3 Germination of seeds ---------- 20
3.4 DNA isolation from needles and megagametophytes ---------- 21
3.5 Environmental dataset ---------- 22
3. 5.1 Aridity index (AI) calculation ---------- 24
3. 5.2 Principal component analyses ---------- 25
3.6 Genotyping data Analyses ---------- 27
3.7 Basic diversity statistics of genotype data ---------- 28
3.8 Population structure analyses ---------- 30
3.9 Outlier detection methods ---------- 31
3.9.1 BayeScan outlier detection method ---------- 31
3.9.2 Bayesian linear mixed model (BayEnv) ---------- 32
3.9.3 Spatial analyses detection (Samβada) ---------- 33

**4. RESULTS** ---------- 36

4.1 Climatic data estimation within mapping population ---------- 36
4.2 Analyses of 384 Illumina GoldenGate genotyping data ---------- 43
4.3 Basic diversity statistics on genotype data ---------- 46
4.4 Population structure estimation ---------- 48
4.5 BayeScan outlier detection method ---------- 52
4.6 Environmental association analyses: Bayesian linear mixed model detection (BayEnv) ---------- 54
4.7 Spatial analyses detection (Samβada) ---------- 61

5. **DISCUSSION** ---------- 62

5.1 The thesis summary ---------- 62
5.2 Limitations of the current study ---------- 64
5.3 Future perspective of the study ---------- 68

**REFERENCES** ---------- 69

**SUPPLEMENTS** ---------- 78
Supplement 1: List of mother trees sampled across the European Alps with geo-reference data provided ---------- 77
Supplement 2: List of SNPs submitted to 384 Illumina GoldenGate genotyping assay ---------- 86
Supplement 3. Assignment of 392 individuals based on the geo-reference data ---------- 98
Supplement 4. List of SNPs failed during 384 Illumina GoldenGate genotyping assay ----------100
Supplement 5. Basic diversity statistics indices on genotyping matrix on polymorphic SNPs ----------103
Supplement 6. (A) Barplot when K=6 for original order for 394 individuals assuming existence of panmixia ----------110
Supplement 7. Fst coefficient estimated as the posterior mean using model averaging (BayeScan) ----------112
Supplement 8: Samβada results for univariate models ----------118

**Parole chiave:** abete rosso, adattamento, genotipizzazione 384 SNPs, analisi degli outiliers

## Riassunto

**Scopo:**
L'obiettivo principale di questo studio è stato quello di investigare sul potenziale genetico adattativo in abete rosso (*Picea abies* (L.) Karst) nell'arco alpino.

**Metodi e Risultati:**
In questo studio, è stato adottato un approccio genomico sul polimorfismo del nucleotide singolo 384 (SNP), secondo il test di genotipizzazione Illumina GoldenGate, per investigare la variabilità genetica e la capacità adattativa in una popolazione di 392 individui di abete rosso campionati nell'arco alpino e georeferenziati. I dati ambientali sono stati ottenuti tramite i dataset WORLDCLIM e IGP MODIS LST satellitare. Per l'analisi statistica multivariata (componenti principali –PCA) è stata applicata su tutte le serie di dati ambientali per ridurre la dimensionalità del campione ed estrarre la più alta percentuale di variazione, per poter procedere poi all'analisi degli outliers.

**Conclusioni:**
La genotipizzazione ha fornito informazioni sul tasso di successo tramite il polimorfismo del singolo nucleotide (SNPs), e ha rivelato fallimento nel caso di 26% SNPs, dove la maggior parte di questi loci sono stati originati dal dataset di Arborea (Canada) su *Picea glauca*. La struttura della popolazione stata stimata utilizzando STRUCTURE (Pritchard *et al*. 2000) per raggruppare potenzialmente gli alberi esaminati in gruppi genetici, sulla base di 394 individui genotipizzati con successo e 280 loci polimorfici. I risultati hanno mostrato l'assenza di una forte struttura della popolazione e questo probabilmente è dovuto ad un *panmixis*. Le analisi degli outiliers sono state eseguite utilizzando un modello bayesiano modello lineare-misto Bayenv 2 (Gunther e Coop, 2013) e il metodo d'analisi spaziale Samβada (Stucki e Joost, 2014). In sintesi, non è stata trovata nessuna forte correlazione tra frequenze alleliche e i gradienti ambientali inclusi nella analisi degli outlier, suggerendo che la selezione nell' abete rosso avviene al livello del singolo locus testato in questo studio.

**Significato ed impatto dello studio:**
Questo studio è uno dei diversi studi esplorativi del potenziale genetico adattativo in abete rosso, realizzati utilizzando un approccio di genomica del paesaggio. I limiti di questo studio risiedono nel numero relativamente ridotto di loci esaminati. Come prospettiva futura, si ritengono essenziali studi di associazione *genome-wide*, così come approcci fenotipo-fenotipo e genotipo-ambiente, in cui almeno 1.000 loci siano considerati. Tali studi sarebbero molto importanti per la migliore conoscenza dell'abete rosso, e costituirebbero un mezzo potente per la ricerca di segnali di selezione e adattamento ai cambiamenti climatici.

**Keywords:** Norway spruce, adaptation, 384 SNPs genotyping, outlier analyses

## Summary

**Aims:**

The main objective of this study was to investigate an adaptive genetic potential in Norway spruce (*Picea abies* (L.) Karst) across the European Alps.

**Methods and Results:**

In this study, we used a genomic approach based on a 384 single nucleotide polymorphism (SNP) Illumina GoldenGate genotyping assay to investigate adaptive genetic variation within a population composed of 392 individuals provided with geo-reference data. An environmental dataset was obtained from WORLDCLIM dataset (Hijmans *et al*. 2005) and PGIS MODIS LST satellite (Neteler, 2010). In addition to temperature and precipitation variables, an aridity index was also included in data analyses. Principal component analyses were applied on all environmental datasets to reduce its dimensionality and extract the highest proportion of variation to be used in outlier detection analyses.

**Conclusions:**

Genotyping analyses provided insight in to the rate of success with single nucleotide polymorphism (SNPs) genotyping and revealed failure in case of 26% SNPs where most of these loci originated from Arborea dataset (Canada) on white spruce (*Picea glauca*). Population structure was estimated by using STRUCTURE (Pritchard *et al*. 2000) to potentially cluster individuals in genetic groups based on 394 successfully genotype individuals and 280 polymorphic loci. The highest delta K value was 2 (Evanno *et al*. 2005) showed there is no strong population structure within the population and suggested of panmixis. Outlier detection analyses were performed using a Bayesian liner mixed model-Bayenv 2 (Gunther and Coop, 2013) and the Spatial analysis method-Samβada (Stucki and Joost, 2014). In the summary, there was no strong correlation between allele frequencies and environmental gradients included in outlier analyses that would suggest selection in Norway spruce at the individual loci tested in this study.

**Significance and Impact of the Study:**

This study was one of several exploratory studies of adaptive genetic potential in Norway spruce by using landscape genomics approaches. Limitations of this study were due to the relatively small number of loci. As a future perspective, genome wide association studies or genotype to phenotype approach as well as genotype to environment when at least 1000 loci included, would be challenging in case of this organism and certainly would provide more power in search for signal of selection.

**Lavori correlati alla Tesi - Papers related to the Thesis.**

- Power and repeatability of outlier and genotype x environment in Norway spruce (*Picea abies* L. Karst)
**Calic I,** Bussotti F, Neale DB (in preparation)

- Characterization of phenotypic traits involved in long-term adaptation to climate change and their genetic base in case of Norway spruce (*Picea abies* L. Karst)
Bonosi L, Mosca E, **Calic I,** Neale DB et al. (in preparation)

- A candidate gene based approach to explore adaptive genetic variation to aridity in Sugar Pine (*Pinus lambertiana*) (in preparation)
Vangestel C, Vazguez-Lobo A, Martinez-Garcia PJ, Wegrzyn JL, **Calic I**, Liechty JD, Neale DB.

- Identifying the genetic basis of partial resistance on Sugar Pine (*Pinus lambertiana*) under white pine blister rust infection (WPBR) (In preparation)
Vazquez-Lobo A, Vangestel C., Martinez-Garcia PJ, Wegrzyn J, **Calic I,** Neale DB

**List of Figures**

Figure 1: Natural distribution map of *Picea abies* across Europe (EUFORGEN, 2013)
Figure 2. Satellite map of Europe with documented drought induced mortality areas
Figure 3: A mapping population consists of 392 trees sampled across the Alps
Figure 4: A 384 Illumina GoldenGate design steps
Figure 5: Germination of *Picea abies* seeds
Figure 6: Germination of *Picea abies* seeds
Figure 7: PGIS Modis LST satellite provides resolution of 250m x 250m
Figure 8: Visualization of correlation matrix between 20 climatic variables and geographic variables
Figure 9: Visualization of loadings for each principal component (seasonal dataset)
Figure 10: Visualization of the cumulative proportion of variance (%) for each principal component
Figure 11: Visualization of loadings for each principal component (monthly dataset)
Figure 12: Visualization of the cumulative proportion of variance (%) for each principal component
Figure 13: Examples of successfully clustered SNP
Figure 14: Examples of unsuccessfully clustered SNPs
Figure 15: Visualization of Wright's Fis inbreeding coefficient
Figure 16. Plot of mean likelihood L(K) and variance per K value from STRUCTURE
Figure 17. (A) The rate of change of the likelihood distribution (mean±SD) and (B) absolute values of the second order change of the likelihood distribution (mean±SD).
Figure 18. Plot of ΔK (Evanno method)
Figure 19: Barplot when K =2 for original order for 394 individuals
Figure 20. R plot visualization of logarithm for the Posterior Odds to base 10 and averaged Fst across populations.
Figure 21. R plot visualization of logarithm for the Posterior Odds to 384 loci under investigation for the signal of selection.
Figure 22. R plots visualization of 280 SNPs and logarithm of Bayes factors for latitude (A); (B) longitude; (C) elevation; (D) PC 1 (the first principal component value); (E) PC 2 (the second principal component value); (F) PC 3 (the third principal component value).
Figure 23: R plots visualizations for correlation between locus id 23 (PGWD1-0634) and logarithm of base 10 of Bayes factors for winter aridity (A) and locus id 96 (PGLM2-0703) and logarithm of base 10 of Bayes factors for autumn precipitation (B).
Figure 24: R plots visualization of 280 SNPs and logarithm of Bayes factors for PC 1 (the first principal component value) (A); (B) PC 2 (the second principal component value); (C) PC 3 (the third principal component); (D) PC 4 (the fourth principal component value.

**List of Tables**

Table 1: Documented cases of drought and heat-induced mortality from Europe
Table 2: Gene products showed to be associated to various climate gradients in different conifer species
Table 3: Mapping population of Norway spruce
Table 4: List of environmental variables.
Table 5: Outputs of *prcomp* function to do principal component analysis in R
Table 6: Summary statistics of the PCA on climatic variables (seasonal dataset)
Table 7: Multivariate measures of climate for the seasonal dataset
Table 8: Summary statistics of the PCA on climatic variables (monthly dataset)
Table 9: Multivariate measures of climate for monthly dataset
Table 10: Basic diversity statistic of genotyping data from 384 Illumina dataset
Table 11: Overview of unsuccessfully genotyped SNPs
Table 12: SNP genotype quality data
Table 13: Basic diversity statistics of genotype data
Table 14: Table output of the Evanno method results
Table 15. Interpretation of Jeffrey's scale of evidence for the selection (Jeffrey, 1961)
Table 16: Summary of SNPs showed to have strong association to geo-climatic variables
Table 17: Norway spruce sampled individuals included for 384 Illumina genotyping assay
Table 18: Summary of candidate genes potentially involved in adaptation process when different outlier detection methods performed
Table 19: The percentage (%) of discovered genes from the total number of genes screened for signal of selection
Table 20: Summary of gene candidates potentially involved in adaptation generated with different outlier methods confirms no repeatability existence

# Acknowledgements

At the end of my time as a PhD student, I am proud of the things I have learnt and accomplished but I am first and foremost grateful to the people who have advised, supported, taught and helped me over the years.

First, I would like to thank to my supervisor **David** for guiding me during three years and giving me a chance to train myself in the best way at UC Davis. It was such a pleasure to be your graduate student. Thank you for keeping always your door open. Your insights into forest population genetics and landscape genomics, the beauty of forest, your hospitality and help during my stay in Davis have all been valuable. **Filippo,** my tutor at UNIFI, thank you for all your support and guidance along doctoral period. I would like to thank to DISPAA Department at UNIFI for providing me with a scholarship and all financial support during my stay abroad. **John Liechty,** I could never mention or thank you enough for all the things you have helped me, big and small, but above all I would like to thank you for always being there as a colleague and never feel like I am bothering you with all my questions. Thank you for introducing me to the world of programming languages, I promise to continue building up my skills in that area! **Pedro Martinez Garcia**, thank you for all your help and for sharing with me with your expertise. **Emily Grau**, thank you for your help with scripts and doing it happily! Thank you for housing help and for your constant cheerful and helpful disposition. Thanks to **Carl Vangestel** and **Alejandra Vazquez Lobo Yuren** for your truly nice collaboration and guidance on Sugar pine projects. I am grateful for consultations and guidance of **Prof. Jeffrey Ross Ibarra** in population genetics. To all people of the Plant Science Department at UC Davis from Neale's group and Ross Ibarra's group, **Randy Famula, Paul Bilinski, Dianne Velasco, Sayuri Tsukahara and Colleen Spurlock**, thank you guys for making a good atmosphere at work, for sharing space and time with me, for laughs and chats, your advices and your friendship!

Thanks to **Lorenzo Bonosi** for the collaboration within PicPhenomics project and for genomic data sharing. **Giuseppe Vendramin**, thank you for opportunity to conduct work in your lab and for all your advices and help. **Ilaria Spanu**, thank you so much for all hard work in the lab, we were such a good team!

Thanks to my family for all their love and support, my Mom and Dad, my grandma, aunt and all near and good people to me.

# 1. INTRODUCTION

## 1.1 *About Conifers*

Conifers is a common name to indicate the members of the order Pinales, (division Pinophyta, class Pinopsida), including approximately 700 species and up to 900 taxa, and represent an ecologically important group (Debreczy and Rácz, 2011). They present a monophyletic group of gymnospermous woody plants with a long evolutionary history dating back to Carboniferous more than 300 million years ago (Farjon and Filler, 2013). Based on fossil records, conifers were distributed worldwide at the beginning of Mesozoic (250 million years ago) when their genetic diversity was high (Stewart and Rothwell, 1993). Conifers remained dominant forests across the world until angiosperms occurred during Early Cretaceous (120 million years ago) facing a competition changes (Farjon and Filler, 2013). Nowadays, conifers are present on all continents except Antarctica and in all climate zones. However, among conifer species, some species of the same genus are only common for a particular small region whereas other species are distributed across large areas or continents. Such case is common for *Picea abies*, which is widely distributed across Eurasia and *Picea omorika*, which is only related to small region of Balkan Peninsula.

Conifers are highly valued for their straight trunks, soft, finely grained and easily workable wood, which all emerged a growth in commercial exploitation of conifer forests worldwide (Debreczy and Racz, 2011). Most of conifer-covered slopes were reduced and finally vanished with only few exceptions remained untouched such as redwoods in California, pine forests at Tasmania's southwest, the alerce forest of the southern Andes (Debreczy and Racz, 2011). However, the importance of conifer ecosystems and forest ecosystems in general has been significantly emerged since they play such an important role in the hydrological cycle, recycling of nutrients and stabilizing of climate.

Within conifers, eight families are recognized where some families are monospecific (Sciadopityaceae) to 231 species numbered in Pinaceae, and from a single country (Japan Sciadopityaceae) to all continents expect Antartica (Cupressaceae). Among all families, the Cupressaceae is the only family with a cosmopolitan distribution (both Northern and Southern Hemisphere). The Pinaceae is the Northern Hemisphere family with the widest distribution, which is shared by several genera but the *Pinus* is making the largest contribution. The Taxaceae is a family of mostly Northern Hemisphere distribution as well as the Cephalotaxaceae with a limited distribution in Eastern part of Asia. The Sciadopityaceae is a family of a single species common just for Japan, which was possible spread from Europe to Japan. The Podocarpaceae is a primarily Southern Hemisphere family with a pan-tropical distribution across Ecuador, Mexico and Japan with genus *Podocarpus* accounting for the most distribution of the family. The Araucariaceae is a family where only three genera make up the family being common for the Southern Hemisphere. The last family, the Phyllocladaceae is also common for Southern Hemisphere of one genus with a limited distribution in Tasmania, New Zealand and Malaysia (Debreczy and Racz, 2011).

## 1.2 *Pinaceae (The Pinus family)*

The family Pinaceae is the largest family in conifers with 240 or more species included widely distributed along Northern Hemisphere with the greatest diversity in temperate to tropical mountain habitats (Debreczy and Racz, 2011). The Pinaceae is coniferous family very important ecologically and of great economic importance as a source of timber, paper pulp, resins and oils (Debreczy and Racz, 2011). The three largest genera, *Pinus* (115+species), *Abies* (55+species) and *Picea* (ca. 35 species) are widely distributed in Eurasia and North and Central America, while *Larix* (10-15 species) has a broad northern-latitude distribution, while *Tsuga* (9 species) and *Pseudotsuga* (4+species) are restricted to portions of Asia and North America (Debreczy and Racz, 2011).

This family dominance is common in conifer forests of the temperate and cold regions. In boreal forest zone are common limited number of species such as *Abies, Larix, Picea* and *Pinus*. In Europe, the family Pinaceae of the boreal forest in Scandinavia is present in most of mountainous regions but absent in lowlands except around the Mediterranean and Black Seas and in the Baltic States, Belarus and Russia (Farjon and Filler, 2013).

The genus *Picea* is represented in Europe with two species, one is widespread, *Picea abies* and the other is a narrow endemic, *Picea omorika*, occurring only in a small area of Bosnia and Herzegovina and a short distance across the border into Serbia along Tara river. The genus *Picea* is accounting for 38 species widely distributed in the northern parts of Northern Hemisphere and presents one of the most dominant conifer genus (Farjon and Filler, 2013). *Picea* is a northerly genus adapted to low or extremely low winter temperatures and growing seasons that can be as short as five or six weeks (Farjon and Filler, 2013).

## 1.3 *Norway spruce natural distribution*

*Picea abies* (L.) H. Karst.1881 is the most widely distributed spruce in Europe, ranging through the alpine systems from the Maritime Alps to the Carpathians, the Balkan Mountains and sporadically in the Dinaric Alps (Debreczy and Rácz, 2011). Across the Europe, its very wide distribution is starting extensively in Scandinavia (Norway, Sweden and Finland) and in northeast part of Europe from Estonia, Belarus and far into Russia. In Central Europe, its presence is limited to the mountainous regions where if present in lowlands is due to its plantations (Figure 1). It occurs from French Hautes Alps to the end of Alps in Austria. It is also present in Germany (the Bavarian Forest, Bohemian Mountains, the Erzgebirge), the Carpathians and the mountains of the Balkans to Montenegro, Macedonia and Bulgaria (Farjon and Filler, 2013). The area of distribution has expanded beyond its natural range, due to its plantations in Belgium, Luxembourg, the Netherlands, in Denmark as well as Great Britain, Ireland and most parts of France (Spiecker, 2000 in Klimo *et al.* 2000).

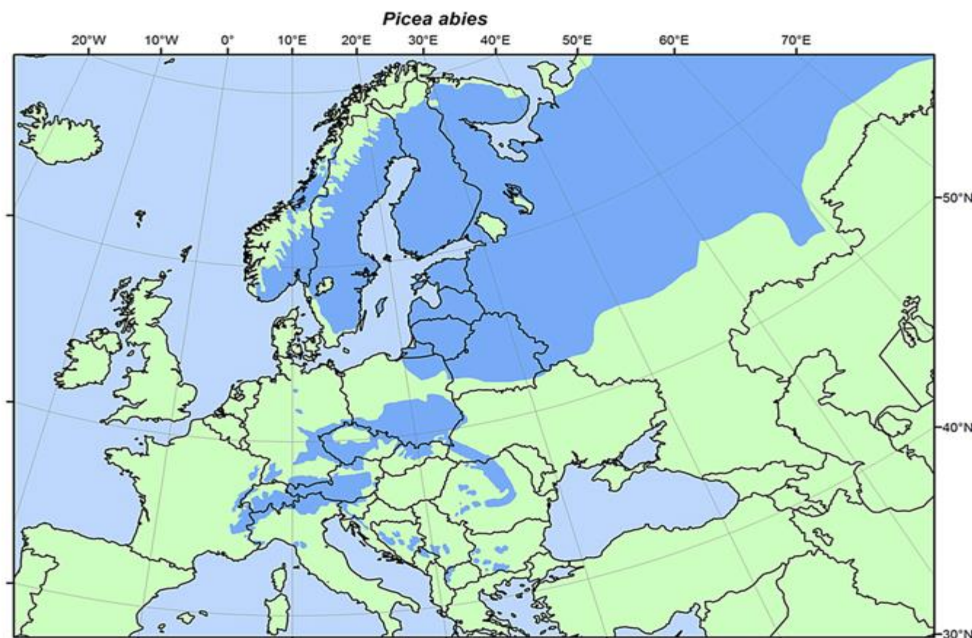



Figure 1. Natural distribution map of *Picea abies* across Europe (EUFORGEN, 2013) (www.euforgen.org)
Norway spruce has migrated from several places of origin causing an expansion of the geographic range primarily to higher mountain regions and lowlands through the interglacial period.

An increase in natural range proportions for Norway spruce within Europe started at the beginning of the 19$^{th}$ century due to the intensive reforestation in order to establish a good source for a timber production (Mullin *et al*., 2011). Nowadays, the range expansion of Norway spruce has been under the human activity influence such as fire, land clearing and extensive logging (Modrzynski, 2007). Norway spruce also has been planted in North America, specifically in eastern Canada (Mullin *et al*., 2011).

## 1.4 *Ecology of Norway spruce*

Two most important climatic factors for the growth of Norway spruce are temperature and precipitation (Nebe, 1968). Environmental factors influence all developmental stages of Norway spruce (Modrzynski, 2007). For instance, amount of solar radiation is one of the important environmental factors important for seed crop abundance in the year prior the flowering since it stimulates bud opening. If environmental conditions are into optimal range, Norway spruce will have a good seed crop every 3-4 years (Modrzynski, 2007). The length of the growing season in Norway spruce depends on photoperiod and temperature. Bud set development is associated to night length, but also other factors such as temperature and moisture requiring the minimum temperature of 0°C for occurrence of deep dormancy. Temperature is important for beginning of the growth for needles and shoots, higher than 5°C, whereas for cambial activity is required temperature above 10°C (Modrzynski, 2007).

The temperature or heat sum requirements of Norway spruce are relatively low, what is evident from the natural geographic range reaching mostly northern and high altitude areas. In the southern parts, Norway spruce reaches high altitude where it forms tree lines or range limits, below which the temperature is too high for the growth (Modrzynski, 2007). Norway spruce is found to be quite tolerant

to low winter temperatures, however it is sensitive to early and late frosts during its growing season (Ellenberg, 1978). It also tolerates high temperatures (up to +46°C) and very low temperature (up to -60°C) at sites where average January temperature of -35°C such as Siberia (Ellenberg, 1978; Schmidt-Vogt, 1977).

Moisture requirements of Norway spruce are quite high, and according to Schmidt-Vogt (1977), Norway spruce is vulnerable to the drought because of its reduced ability to uptake water when it is scarce. Regarding soil requirements, Norway spruce is able to develops optimally on soils that are mesic, loamy and medium rich in nutrients with an optimal pH range lies between 5.3 and 6.0, however it can grow on soils with pH values ranging from 3.4 to 6.7 (Obminski, 1977).

Most disturbances in Norway spruce stands are caused by both biotic and abiotic factors, such as storms, snow and insects. Drought, frost, fungal diseases and herbivores are causes of major disturbance in Norway spruce stands (Schmidt-Vogt, 1989). Norway spruce tolerance to low temperature is well entertained if properly conditioned in late summer and hardened in fall, being able to tolerate to about -60°C during winter dormancy (Schmidt-Vogt, 1977). On the other hand, as it was reported by Christersson and Fircks (1990) observed seedling injury was up to 100) at temperature -4°C. Harmful impact of high temperatures is common for your seedlings when exposed at the soil surface (up to 65°C). Such temperatures cause protein denaturation and lead to seedling mortality, whereas a harmful effect of high temperature on mature trees rarely occurs (Schmidt-Vogt, 1977). The tolerance to heat stress is triggered through dehydration in tissues and heat shock proteins synthesis (Modrzynski, 2007). Drought induced mortality in Norway spruce was observed in nurseries and young stands, where in old stands drought damages are minor (Modrzynski, 2007). If present, drought changes are causing physiological changes such as decrease in gas exchange rates and peroxidase activity and carbohydrate exudation from root system (Schmidt-Vogt, 1977).

## 1.5 *Climate change and its impact on conifers*

Forest ecosystems are being rapidly and directly transformed by the land uses due to the expanding human populations and economies (Allen *et al*. 2010). Less evidence is available nowadays about impact climate change could have on the world's forest. One of the most important challenges recognized globally is the climate change effect on ecosystems (Koskela *et al*. 2007). Although widely acknowledged, the global temperature has raised up (0,5°C since 1970) due to the increased emissions of greenhouse gases, future predictions are indicating further increases in mean temperature (2-4°C globally) with significant mortality issues in some parts of the world (Christensen *et al*. 2007; Seager *et al*. 2007) as well as extreme droughts or high temperature raise (Allen *et al.* 2010). Average temperatures are expected to increase by 2-4°C in Europe over the next fifty years causing significant changes in regional and seasonal patterns of precipitation (IPCC 2007b), creating novel environmental conditions to which forest ecosystems will be exposed.

Forest ecosystems might not be able to support novel climate changes due to their long life span and time span of the climate change (Kremer, 2007). Based on the palaeoecological data assumptions, the rate of evolutionary change is much slower than the rate of present climate change (Davis *et al*. 2005), leading to the fact that forest species as a sedentary, long lived organisms are under the risk of the potential dieback caused by global climate changes. Trees belong to ecologically and economically important group of the terrestrial ecosystems where these novel changes are about to occur.

Plenty of evidences of die-off cases due to drought and high temperatures on conifer species are already available and reported across the world (Table 1). In Europe, mortality cases due to dry and warm conditions in the 1990s and 2000s across Mediterranean regions, including death among woody species in Spain such as *Pinus sylvestris* (Penuelas *et al*. 2001; Martinez-Vilalta and Pinol, 2002) and in increases in mortality of *Pinus sylvestris* in Switzerland and Italy (Dobbertin and Rigling, 2006, Bigler *et al.* 2006) (Figure 2). For Norway spruce, has been reported a case of dieback due to summer drought paired with biotic stressors in Norway (Solberg, 2004) and in Switzerland due to drought and high summer temperatures in 2003 correlated to bark beetles (*Ips typographus*) reported by Forster *et al.* (2008).

Climate induced mortality of conifer species have been reported in other parts of the world as well. For instance, in North America has been reported drought induced mortality on example of *Pinus contorta* in British Columbia (Kurz *et al.* 2008a) and several spruce species in Alaska (Berg *et al*. 2006).

Figure 2. Satellite map of Europe with documented drought induced mortality areas (Table 1). Right picture presents mortality of *Pinus sylvestris* in Switzerland (1999), By: Wermelinger B. Left picture presents die-offs of *Pinus sylvestris* in Spain (2006), By: Navaroo-Cerrillo R. (References in Allen *et al*. 2010)

Many recent examples of drought and heat-induced mortality from around the world suggest that no forest type is invulnerable to climate change (Allen *et al*. 2010). Koski *et al.* (1997) reported populations of Norway spruce are facing severe threats in some regions in Central Europe. Populations are in danger of possible dieback due to the industrial air pollution and global warming. Although it is common high levels of intrapopulation genetic variation and known high adaptability of Norway spruce, decline and mortality rates are predicted to occur (Koski *et al*. 1997).

Table 1: Documented cases of drought and heat-induced mortality from Europe (1970 to present), ID numbers refers to locations mapped (Allen *et al*. 2009).

| Location | Years of mortality | Forest type | Dominant tree | Climate anomaly linked to mortality |
|---|---|---|---|---|
| Europe (Western Central) ID: 2 | 1970-1985 | Temperate conifer and broadleaf | *Abies* spp. *Picea* spp. *Pinus* spp. | Repeated droughts |
| Italy (South Tyrol) ID: 6 | 1992 | Temperate mixed conifer | *Pinus sylvestris* | Multi year drought |
| Austria (Tyrol) ID: 8 | 1991-1997 | Temperate mixed conifer | *Pinus sylvestris* | Seasonal droughts |
| Italy (Aosta) ID: 9 | 1985-1998 | Temperate mixed conifer and broadleaf | *Pinus sylvestris* | Multi year drought |
| Norway ID: 12 | 1992-2000 | Temperate conifer | *Picea abies* | Multi year summer drought |
| Switzerland ID: 16 | 2003 | Temperate conifer and broadleaf | *Picea abies* | Drought, high temperatures |
| Spain ID: 19 | 2004-2006 | Temperate conifer | *Pinus sylvestris*, *Pinus nigra* | Multi year drought |
| France ID: 23 | 2003-2008 | Temperate mixed conifer and broadleaf | *Quercus* spp. *Abies* spp. *Picea abies* *Pinus* spp. | Spring, summer droughts Scorching heat |

## 1.6 *Phenomena of adaptation*

Adaptation, or the hereditary adjustment to the environment, is one of the universal features of life (Grant, 1963). An adaptation could be defined as a characteristic evolved by a natural selection, which enchases the survival or reproduction of organisms (Futuyma, 2005). With the climate changes causing environment changes, optimal sets of adaptations with maximizing fitness under local conditions might shift accordingly (Kremer *et al.* 2012). Aitken *et al.* (2008) described three possible scenarios for forest ecosystem how to cope with novel environmental changes as (1) migration to track ecological niches spatially; (2) adaptation to new conditions in current location and (3) extirpation. Existence of phenotypic plasticity and effective genotypes enables forest species to tolerate the environment change (Savolainen *et al*. 2007).

Trees as long-lived sedentary organisms with a capacity of long distance dispersal via seed or pollen are not estimated enough about their genetic basis of adaptation potential given an intensity and directionality of climatic change (Kremer *et al*. 2012). However, the strict criterion for a local adaptation is that one population has higher fitness at its native site than any other population at the same site (Savolainen *et al*. 2013). At other sites, populations would have lower fitness, however that would not mean locally adapted population would grow better elsewhere but only their growth is the best at site where locally adapted (Kawecki and Ebert, 2004).

Understanding the genetic basis of adaptation is possible though the way of estimation distribution of allele frequencies among populations. Evolutionary forces that drive allele frequency distribution changes across time are mutation, gene flow (migration), genetic drift (random changes in allele frequencies from generation to generation due to sampling effects), natural selection and mating system (the degree to which sexual reproduction occurs through self-pollination, consanguineous mating between related or unrelated individuals) (Aitken, 2004). Gene flow influences the process of adaptation via seed migration or pollen flow when new areas are colonized (Austerlitz *et al*. 2000). The potential of forest trees to adaptation is reflected through gene flow (mostly by pollen) and natural selection (Savolainen *et al*. 2014). On the other hand, genetic drift is more reflected than other evolutionary forces in small or isolated populations or species with small ranges and experienced major bottlenecks (Aitken, 2004). The adaptive potential one population could have depends on phenotypic variation, strength of selection, fecundity, interspecific competition and biotic interactions. Tree populations from boreal and temperate zones show moderate to strong clines in phenology and growth along temperate gradients indicating substantial adaptation (Aitken *et al*. 2008).

Trees as long-lived organisms may not have evolutionary rates as high as short generation organisms (Smith and Donoghue, 2008) raising the concern about having an exact match within evolutionary change with rapid climate change. Trees experienced strong evolutionary responses during glacial and interglacial periods when dramatic environmental changes occurred leading to their replacement to more favorable habitats but also their extinction (Kremer, 2010). Based on fossil records, some plant species that previously existed in Europe end of the Tertiary period nowadays are present in North America, however there was no tree extinction in recent period, indicating that trees have developed ability to migrate or adapt to environmental change (Kremer, 2010). Capacity for colonizing new areas for some forest tree species are predicted to be limited, such as Scots pine, which is predicted to decline (Kellomaki *et al*. 2001).

Savolainen *et al*. (2013) described three possible population genetics approaches for detecting the footprints of local adaptation at the molecular level: population differentiation through scans of the Wright fixation index (Fst), correlation between allele frequencies and environmental variables and Fst-Qst comparisons. Common garden experiments proved to been informative in estimation of adaptive genetic potential, however their informative power depends on the number of different test environments, the duration of tests and the genetic diversity of experimental populations (Neale and Kremer, 2011). With discovery of genomic data, genome scans became an alternative to common garden experiments, as time consuming, costly and logistically challenging approaches. However, through combination of genomic and more traditional approaches, intensive efforts have been made to identify the most ecologically important traits and their underlying genes (Neale and Kremer, 2011). Although approximately 290 genes have been investigated for evidence of balancing or directional selection, around 20% (55 genes) showed departures from neutrality (Neale and Ingvarsson, 2008). Population genetic analyses provide enriched exploration of adaptation signatures based on sampled DNA sequences from multiple individuals especially when common garden or phenotypic selection, are not feasible. However, Tiffin and Ross Ibarra (2014) emphasized the usage of population genetic analyses that provides enriched exploration of adaptation based on sampled DNA sequences from multiple individuals also when common garden experiments or phenotypic selection analyses are not feasible.

## 1.7 Adaptive genetic variation in conifers with emphasis on Norway spruce (Common garden trials)

Evidence of evolutionary responses through local adaptation in trees may originate from two sources: provenance trials and patterns of clinal variation (Savolainen *et al*. 2007). Selection along environmental gradients in continuous populations often results in clines. Development of the genetic variation that provides increase of phenotypic clines depends on balance between selection and the average disperse distance of the organism (Savolainen *et al*. 2007). Generally, common garden experiments provided an insight in phenotypic proportion of variation between traits and abiotic and biotic environmental factors (Savolainen *et al*. 2007).

Traits related to adaptation to specific environmental stresses could be assessed through long-term field provenance trials such as phenology (e.g. timing of bud burst, bud set, leaf abscission, pollination and seed maturation); cold-hardiness and drought-related traits. However, adaptive traits are more commonly assessed in short-term nursery, growth chamber or greenhouse in more controlled conditions (Aitken, 2004). Long-term provenance trials enabled an opportunity to study differences among populations from different environments as well as effects of a changing environment on these populations by substituting spatial for temporal environmental variation (Aitken 2004).

Common garden plantations have provided an insight about variability in quantitative adaptive traits among and within wild populations of forest tree species (Skroppa and Johnsen, 2000). The analyses of phenotypic clinal variation due to the selection can be a productive approach to uncover local adaptations and their genetic basis (Savolainen *et al*. 2007). Multiple-site provenance trials (also known as common garden plantations) could serve to examine plastic responses of populations in novel environment when planted in sites where environmental gradients are different (Alberto *et al*. 2013).

A number of traits known to have adaptive significance have been observed through common garden studies. In the previous conducted provenance trials, a substantial variation has been observed among populations for traits related to the climatic adaptation (Skroppa and Johnsen, 2000). One of the examples of such study is trial conducted on one-year-old seedlings of Norway spruce, when a clinal variation among populations as observed and the latitude and altitude of the natural stands explained 88% of the variability (Daehlen *et al*. 1995). Skroppa and Johnsen, 2000 described several cases of clinal variation for different traits on case of Norway spruce such as cessation of shoot elongation at the age of six, annual growth rhythm traits, duration of shoot growth, and basic wood density, pointing out that that higher percentage of variation has been observed on population level rather than on family level and it could be explained by different mechanisms influence the genetic variability within and among populations (Skroppa and Johnsen, 2000).

Phenotypic clines along environmental gradients are common in forest trees such as variation in the mean timing for bud set in *Pinus sylvestris* along latitudinal gradient (Mikola, 1982). Beside bud set, other traits have been extensively studied such as phenology in general and flowering time for which strong candidate genes are available and belong mostly to the photoperiodic pathway including circadian clock (Gyllenstrand *et al*. 2007).

In study conducted by Chen *et al.* (2012) was analyzed genetic and phenotypic variation in bud set in Norway spruce, a highly heritable adaptive trait, for which variation appeared to be strongly clinal. Chen *et al*. (2012) emphasized that Norway spruce populations, just as in the case of other forest species, are strongly differentiated for bud set indicating very strong phenotypic differentiation for this particular trait. Additionally, Chen *et al.* (2012) analyzed single-nucleotide polymorphism (SNP) frequencies at candidate genes for timing of bud set in 18 populations of Norway spruce, providing capture of the clinal variation at candidate gene SNPs correlated with latitude. The clinal variation has been finally confirmed in gene expression study on subset of genes for which clinal variation was detected for SNP allele frequencies (Chen *et al*. 2012).

## 1.8 Genotype versus environment in conifers with emphasis on Norway spruce

Landscape genomics integrates both landscape ecology and population genetics with an aim to investigate how landscape elements (such as forest) and environmental factors influence the spatial distribution of genetic variation. One of the main research questions of landscape genetics is how environmental factors such as temperature or precipitation affect adaptive genetic variation due to the climate change (Holderegger *et al*. 2010). As genomic data became more available nowadays, genome scans for signatures of adaptation became more common (Tiffin and Ross Ibarra, 2014). In genomic era, the first approach was population genomics such as Fst outlier method, which was common for about decade (Holderegger *et al*. 2010). Landscape genetics became available more recently and it directly use environmental variable to highlight molecular markers linked to genomic regions, which are under the selection (Holderegger, 2008). In particular, allele frequencies in individuals either in the populations are correlated to any of environmental variables such as: temperature, precipitation, aridity, altitude etc. (Holderegger, 2008). Genetic variation is often influenced by latitudinal or altitudinal clines, however climate is the most important driver of adaptive phenotypic traits (Eckert *et al*., 2010a).

Several landscape genomics studies on the conifer species have been conducted in searching for an estimation of adaptive genetic variation. One of the first attempts on non-model organism to detect outlier loci under the selection on Norway spruce was done by Scalfi *et al*. (2014). In this study was reported usage of *F*st outlier detection analyses in a combination with regression analyses. In this study were used 384 SNPs representing 290 genes used for genotyping at micro-geographic study (two altitudinal-transects in the Italian Alps) for a total of 38 populations and macro-geographic study across 27 populations belonging to the Alpine and Hercyno-Carpathian domains. As result, seven loci were detected with Fst outlier detection; six other outliers were detected through population clustering (Bayesian simulation and hierarchical island model) and 15 outliers with regression analyses.

Another approach was described by Chen *et al*. (2012) when analyzed genetic and phenotypic variation in bud set, a highly heritable and adaptive trait, among 18 populations of Norway spruce for correlations with latitude. In this study were included genotypes for 137 single-nucleotide polymorphisms (SNPs) selected from 18 candidate genes putatively affecting bud set and 308 control SNPs selected from 264 random genes for estimation of patterns of genetic structure and environment.

Several methods have been applied to analyze clinal variation such as linear regression on latitude assuming that geographic variation is associated to an environmental gradient covarying with latitude as well as Regression Monte Carlo simulation to investigate to what extent a significant clinal variation observed at one nucleotide site could be due to LD with another SNP (Berry and Kreitman, 1993). Bayesian generalized linear mixed model (Bayenv) analyses was applied to detect clinal variation for SNP allele frequencies in correlation to environmental variables. The last method used in this study, was BayeScan method (Foll and Gaggiotti, 2008). Chen *et al*. (2012) proved that variation in bud set is clinal one. As result of analysis of clinal variation in allele frequencies, 28 significant SNPs including 11 from candidate genes were detected.

Di Pierro *et al*. (in preparartion) reported approach on Norway spruce to determine and quantify patterns of adaptive response in natural populations across the Italian Alps and Apennines. Across the Italian species range, 24 natural populations were sampled. Genotyping was done for 384 selected Single Nucleotide Polymorphisms (SNPs) from 285 genes. As results of this study, were revealed five Fst outliers and seven SNPs associated to one or more climatic variable, precipitation more than temperature (Di Pierro *et al*. in preparation).

Investigation of local adaptation potential has been performed also on other conifer species. One of examples is the study on four different conifer species: *Abies alba*, *Larix decidua*, *Pinus cembra* L. and *Pinus mugo* Turra across the European Alps is by Mosca *et al*. (2012). As the result, six SNPs were detected in *Larix decidua* and 18 in *Pinus mugo* Turra, which were associated to PC1 (the first principal component score) corresponding to winter precipitation and seasonal minimum temperature. In *Abies alba*, four SNPs were associated with PC2 (the second principal component score) corresponded to the seasonal minimum temperature. In all species, except *Abies alba,* was present a strong effect of seasonal temperature and precipitation when the strongest signal of correlation originated from association to seasonal minimum temperature for *A. alba, P. cembra and P. mugo* (Mosca *et al*. 2012).
One of the first attempts to illustrate population genomics approach on non-model organism such as loblolly pine (*Pinus taeda*) was conducted by Eckert *et al*. 2010. Loblolly pine is a species distributed across the southern eastern part of USA, ranging from Texas to Delaware enabled to determine genetic variation among natural populations based on isozymes and nuclear microsatellites (Eckert *et al*. 2010). In this study was presented a genome-wide dataset of single nucleotide polymorphisms (SNPs) typed from 1730 loci in a mapping population composed of 682 loblolly pine individuals. Totally were examined 54 populations for allele correlations to multivariate environmental variables (the principal component scores) (Eckert *et al*. 2010). Very strong association between allele frequencies and geo-climatic multivariate variables (the principal components) was common for 22 SNPs, whereas totally 48 SNPs showed the strongest signal of correlation (i.e. Bayes Factors (BFs.) > 100) describing temperature, growing degree days above 5C, winter aridity, overall aridity, precipitation as well as aridity during summer and winter.

## 1.9 Gene groups associated to environment among conifers

Mosca *et al*. (2012) reported several associated SNPs located in genes encoding proteins affecting plant response to abiotic stress. Loci associated to multivariate components (PCs) were found in genes involved in lignin biosynthesis. Mosca *et al*. (2012) observed 3 loci where one was detected in *Pinus cembra* and other two loci were detected in *Pinus mugo*. Additionally, two loci encoding for heat-shock proteins (HSP) were detected in *Abies alba* (one) and in *Larix decidua* (one) and they were also correlated to PCs scores. Several associated SNPs were located in genes encoding proteins involved in translocation mechanisms, two loci in *Larix decidua*, two loci in *Pinus cembra* and one on *Pinus mugo*. Overall, seven associated SNPs were located in genes common among all four species, when these loci were mainly involved in metabolic and oxidation-reduction processes and in regulation of stomata movement. Among climatic variables, the strongest correlation in this study was related to the seasonal minimum temperature and winter /autumn precipitation (Mosca *et al*. 2012).

Di Pierro *et al*. (unpublished) reported seven associated loci to environmental variables, where for four loci was known putative function and three others were anonymous. Known putative functions were: encoding for SNF2 protein family; encoding for a putative conserved domain: oxygen evolving enhancer protein1 (involved in photosystem II stabilization), encoding for a protein similar to phosphoenolpyruvate carboxykinase 2 (critical enzyme involved in the gluconeogenesis) and locus encoding for a sequence with a high similarity with NAD(P)-linked oxidoreductase-like protein in *Arabidopsis thaliana* (involved in oxidation reduction on chloroplast thylakoid membrane). Di Pierro *et al*. (unpublished) observed more correlation to precipitation than to temperature among climatic variables included in outlier detection analyses.

Scalfi *et al*. (unpublished) attempted detection of loci with different putative functions both on macro and micro geographic scale. As summary of Fst outlier detection methods, only at macro-geographic scale have been detected seven outlier loci, where one has a similarity with protein of *Zea mays* (ACG24319) and another was similar to sucrose synthase of *Pinus halepensis* (AY705802.1). Other detected loci were involved in various biological processes such as: (a) transcription factor, (b) translation-elongation factor, (c) UBX domain-containing protein, (d) acyl-CoA thioesterase or (e) acetyltransferase component. Totally seven outliers were detected with Fst approach and one locus in detected in a gene involved in plant defense and other two in lipid metabolism. Since at macro-geographic scale, analyses were conducted twice, loci detected using populations assigned according to their geographic position was only one and located in *Picea glauca* (gb|BT10728.1|) and encoding for a nucleic acid-binding protein in *Arabidopsis thaliana*. Considering another approach, clustering based on STRUCTURE (Pritchard *et al*., 2000), two loci detected being involved on a gene encoding for a sucrose synthase in *Pinus halepensis* while another locus detected was located in gene encoding for a F-box protein. In this study, regression analyses were also performed. At micro-geographic scale analyses only two loci showed significant correlation with climatic variables such as temperature and precipitation where one locus has similarity with the pentatricopeptide (PPR) repeat containing protein of *A. thaliana*. In the macro-geographic scale analyses, totally 38 loci from 227 tested were significant, where only 12 loci were significant.

In study conducted by Eckert *et al*. (2010a), more than half of identified loci were located in genes that code for proteins with unknown function (hypothetical proteins) referring to that time lack of

sequenced loblolly pine genome but referring to *Arabidopsis thaliana* reference genome or lack of sequence similarity to genes in model plants. Most of other identified loci are related to plant response mechanisms to abiotic stress such as calcium-dependent protein kinases (CPDKs) which approved to be a key signal to transduction of osmotic stress signaling, short-chain dehydrogenase/reductases and TIFY domains. Proteins with TIFY domains are known to respond to a wide range of abiotic stresses such as drought, salinity and low temperature. Eckert *et al*. (2010a) described other potentially adaptive genes encoding for proteins responsible for oxidative stress (e.g. oxido-reductases, peroxidases, thioredoxin-like proteins); cell membrane related proteins (e.g. nodulin, $K^+$: $K^+$ antiporter) and genes encoding for sugar metabolism (e.g. trehalose-6-phosphate phosphatases, TPPs). Additionally, Eckert *et al*. (2010b) reported one SNP located in gene with known gene function that is correlated to ubiquitin-specific protease, proteins quite abundant in eukaryotes and being expressed when ongoing increased protein degradation as response to environmental stress.

Furthermore, Eckert *et al*. (2010a) reported five more loci correlated to aridity where as described; these loci were primarily involved with abiotic stress response to temperature and drought. Five loci showing significant association within genotypes and aridity gradients were located in genes encoding for following: Hexose:hydrogen symporter, Photosystem II protein, CSHC4-type RING finger, MATE efflux family protein and UDP-galactose transporter.
Grivet *et al*. (2011) examined environmental associations between variation at candidate genes and climatic variables with an aim to identify genes, which may be targeted by selection on two Mediterranean conifer species: *Pinus pinaster* Ait. and *Pinus halpenses* Mill. In this study was assessed the impact of natural selection on the same set of candidate genes related to drought tolerance. As result, environmental association revealed loci correlated to temperature, when one of these loci was common in both pine species. Genes discovered belonged to dehydrins, which present a small group of multigene family of intracellular stabilizers that plays a major role in cell protection against desiccation. Additionally, *4cl* gene family has been detected by environmental association analyses. This gene family is involved in the production of basic enzymes of the phenylpropanoid metabolism that present important metabolites acting against biotic and abiotic stresses (Rani *et al*. 2009) as well as encoding key enzymes in biosynthesis of lignin and several studies have demonstrated their involvement in the plant growth (Wagner *et al*. 2009).

Prunier *et al*. (2011) scanned the genome of the boreal conifer black spruce (*Picea mariana* [Mill.] B.S.P.), a dominant tree of the North American boreal forest, for gene single-nucleotide polymorphisms (SNPs) potentially involved in adaptations to temperature and precipitation variations. In this study were included 583 SNPs from 313 genes potentially playing adaptive roles. Prunier *et al*. (2011) detected 14 outlier SNPs correlated to temperature and 16 to precipitation. Loci were located in genes with different putative functions where these correlated to temperature are belonging to C2H2 zinc finger family; R2-R3 MYB, Zinc-binding family, C3HC4 RING finger, AP2, WD-40 repeat family, MBF, Peroxidase, Glycosyl-hydrolase. Loci found to be correlated to precipitation belonged to B-box zinc finger, C3HC4 RING finger, C2H2 zinc finger, Zinc-binding family, Mov31 family, Ubiquitin, LEA, NAC, HD-Zip etc.
De La Torre *et al*. (2014) performed a genome-wide association study on white spruce [*Picea glauca* (Moench) Voss] and Engel-mann spruce [*Picea engelmannii* Parry], two conifer species common in western North America that hybridize extensively in British Columbia and the western part of Alberta (Canada). This study aimed to assess levels of admixture and introgression and to identify loci putatively

involved in adaptive differences between species by using a panel of 311 candidate gene single nucleotide polymorphisms (SNP) from 290 genes. The main objective of this study was to determine the glacial and postglacial re-colonization patterns between these widely hybridizing conifer species. As result, two species are showing to have long history of hybridization and introgression whose integrity has been maintained by a combination of strong environmental selection and reduced current interspecific gene flow (De La Torre *et al.* 2014). Twenty loci showed evidence of divergent selection where six loci were Fst outliers and associated to climatic gradients, whereas other fourteen loci were either outliers or showed association to climate. These loci were located in genes involved in carbohydrate metabolism, signal transduction and transcription factors (De La Torrre *et al*. 2014).

Overall, several group of genes associated to climatic gradients have been identified in different conifer species. Based on the literature sources available so far, seven gene groups were shared among broad conifers, indicating the repeatability in genes also when different outlier detection approach used. The most common gene products are listed as following: (a) lignin biosynthesis present; (b) sugar metabolism; (c) peroxidases; (d) photosystem II protein; (e) Photosystem II protein, (f) Glycosyl hydrolase (g) RING proteins and (g) ubiquitin. (Table 2).

Table 2: Gene products showed to be associated to various climate gradients in different conifer species

| Phenotype | Conifer species | Candidate gene | Reference |
|---|---|---|---|
| Drought | *Pinus pinaster* | Lignin biosynthesis | Grivet *et al*. (2011) |
| Temperature, precipitation | *Pinus cembra* *Pinus mugo* | | Mosca *et al.* (2012) |
| Aridity | *Pinus taeda* | Sugar metabolism | Eckert *et al*. (2012b) |
| Temperature, precipitation | *Picea abies* | | Di Pierro *et al*. (unpublished) |
| Aridity | *Pinus taeda* | Peroxidases (WF, AS) | Eckert *et al*. (2012b) |
| Temperature, precipitation | *Picea mariana* | | Prunier *et al*. (2011) |
| Aridity | *Pinus taeda* | Photosystem II protein | Eckert *et al*. (2012a) |
| Temperature, precipitation | *Picea abies* | | Di Pierro *et al*. (unpublished) |
| Temperature, precipitation | *Picea mariana* | Glycosyl hydrolase (translocation mechanism) | Prunier *et al*. (2011) |
| | *Picea glauca* *P. engelmanii* | | De La Torre *et al*. (2014) |
| Aridity | *Pinus taeda* | RING proteins | Eckert *et al. (*2012a) |
| Temperature, precipitation | *Picea mariana* | | Prunier *et al*. (2011) |
| Aridity | *Pinus taeda* | Ubiquitin | Eckert *et al*. (2012b) |
| Temperature, precipitation | *Picea mariana* | | Prunier *et al*. (2011) |

## 2. RESEARCH OBJECTIVES

The main objective of this study is to estimate an adaptive genetic potential within mapping population of Norway spruce across European Alps for a climate change. It is already known that the most important environmental gradients, which influence local adaptation among forest trees, are temperature, precipitation and aridity. Current climatic predictions suggest temperature could rise 1.5-5.8°C in the next hundred years significantly increasing dieback of forest species across the globe. Norway spruce is known to be sensitive to increases in temperature and droughts in temperate and boreal conifer forests. Many recent examples of drought and heat-induced mortality cases across the world, strongly suggest that there is no forest type or climate zone invulnerable to climate change. How rapid changing climatic change will affect Norway spruce growth remains unclear, since the genetic basis of adaptation is not well understood.

This study provides an insight to adaptive genetic variation within 392 geo-referenced individuals sampled across European Alps. The research question in this study is to identify presence of adaptive loci across the genome of Norway spruce. Herewith, we applied an approach where environmental data are directly associated to genomic data of collected samples.

The first research objective was to determine genetic diversity of Norway spruce by genotyping 384 single nucleotide polymorphism (SNPs) from natural populations. The second research objective was to estimate the correlation between climatic variables (e.g. temperature, precipitation or aridity) and genetic variation. Several outlier detection methods provided an insight of correlation between allele frequency and environmental gradients. The last research objective was to discover possible putative genes and responses in these genes underlying responses to climate.

**3. MATERIAL AND METHODS**

## 3.1 *Sample collection*

Needle tissue and seeds were collected in winter 2011/12 across Italian sites (Bonosi L. and Ghelardini L. FEM, San Michele) whereas the rest of sampling material was obtained from cooperating institutes (WSL Birmensdorf; BFW Vienna; INRA Avignon; Grozdarski Institut Slovenije). 392 mother trees originating from three main provenances Italy-Slovenia-France, Austria and Switzerland constitute a population of Norway spruce (Figure 3). Each mother tree was provided with geo-reference data (latitude, longitude and altitude) except for Austrian provenance where individuals were assigned randomly on Google maps (Bonosi L.) (Supplement 1).

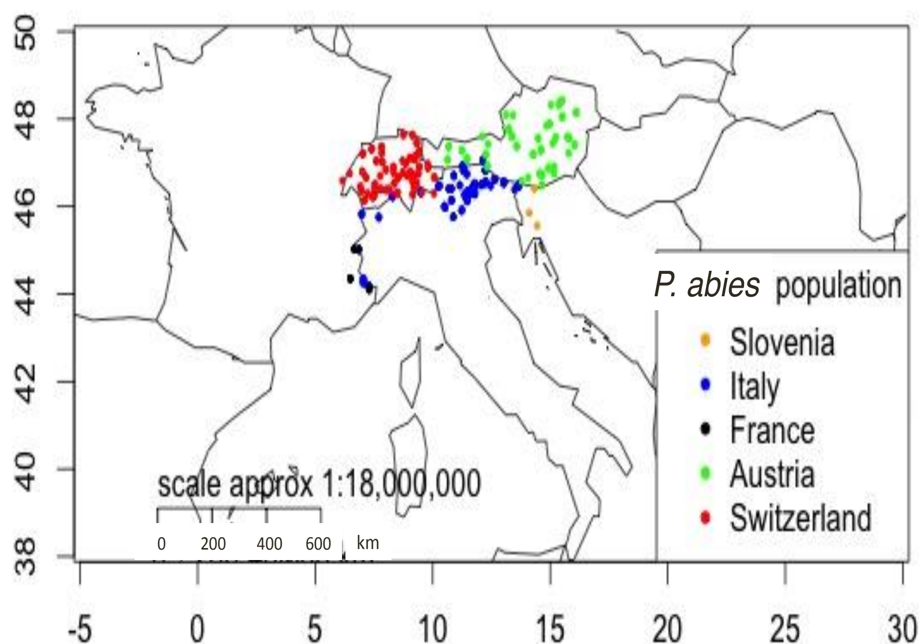



Figure 3: A population consists of 392 trees sampled across the Alps

Each mother tree is geo-referenced by location (site) ($n_{sites}$=152). An average number of sampled trees per site is 3~4 trees/site (Italy), 2 trees/site (Austria), 2 trees/site (Switzerland), 4 trees/site (France) and one tree/site (Slovenia) (Table 3). Needle tissue from 196 individuals and seeds from 187 individuals were sealed in aplastic bag and stored at -20°C until further processing for DNA isolation.

All mother trees were sampled at isolated points across Alps. An average elevation at which trees were sampled was 1066.33m above Earth's sea level (Slovenia); 1433.07m above Earth's sea level (Italy), 1624.5m above Earth's sea level (France); 1036.5 above Earth's sea level (Austria) and 1254.05m above Earth's sea level (Switzerland).

Table 3: The population of Norway spruce

| Country | No. trees | No. sites | Tissue sample | Sample source |
|---|---|---|---|---|
| Slovenia | 3 | 3 | needles | Godzarski Institut Slovenie |
| Italy | 190 | 53 | needles | Bonosi, Ghelardini(FEM) |
| France | 12 | 3 | needles | INRA Avignon, France |
| Austria | 80 | 37 | mega's | BLW Vienna |
| Switzerland | 107 | 56 | mega's | WLS Birmensdorf |

## 3.2 *Illumina 384 GoldenGate chip design*

SNP genotyping (single nucleotide polymorphism) Illumina GoldenGate 384 chip was designed by integrating three different SNPs sources.

The first set of SNPs is from Comparative Re-sequencing in Pinaceae (CRSP) project (http://dendrome.ucdavis.edu/NealeLab/crsp/). CRSP project was based on Sanger re-sequencing of a panel of 12 unrelated Norway spruce trees (haploid megagamethophyte DNA) using subset of 1024 primer pairs derived from almost 1000 loblolly pine expressing tags (ESTs) including genes involved into several biological mechanisms (Scalfi *et al*. 2014 and Di Pierro *et al*. in preparation). SNPs performed well in the previous studies based on the successful genotyping and quality, in particular a subset of 153 SNPs derived from Scalfi *et al*. 2014 and Di Pierro *et al*. (in preparation), were used for a 384 chip design. The final score of SNPs from CRSP project annotated for different biological processes is 94. (see Figure 4. for details).

Additionally, a second set of SNPs delivered from sequencing panel of Evolutionary Biology Center at Uppsala (Sweden) has been used. From the total set of 1146 SNPs, a final set of 102 SNPs including 24 of candidate genes, were selected for the genotyping. The source of these SNPs is from re-sequencing panel for discovery of new SNPs of 48 Norway spruce alleles. These SNPs were also applied in the study of Chen *et al.* (2012) on Norway spruce and they were putatively involved into photoperiodic pathway, shoot apical development and circadian clock.

Finally, the remaining set of 374 SNPs is from Arborea project on *Picea glauca* (white spruce) Canada (http://www.arborea.ulaval.ca/). These SNPs were generated from 1964 SNPs from 1485 genes, which were initially identified and successfully genotyped in *Picea glauca* (white spruce) and used to be tested in 12 (diploid) individuals of Norway spruce from 12 populations in central Europe (Latvia, Poland and Germany). These 1485 genes were transcription factors either candidate genes related to growth, wood formation, cell wall, lignin synthesis (Chen *et al*. 2012). Testing was done by submitting SNP design panel to Illumina GoldenGate SNP genotyping resulting in discovery of 384 valid SNPs from 340 genes. A group of 374 SNPs are obtained in the collaboration with EB Uppsala, Sweden (Dr. Martin Lascoux) was used for selecting SNPs and the final number of SNPs selected was 188 SNPs originating from genes

putatively involved into photoperiodic pathways and the rest of control SNPs being involved into different biological functions.

The process is initiated by selecting and submitting a list of requested loci to Illumina genotyping center (Parco Tecnologico Padano di Lodi, Italy). Upon submission, Illumina evaluated the list with Assay Design Tool (ADT) to ensure successful assay development. ADT tool was the first step to create an assay panel including these loci that have a high likelihood of success for genetic analyses performance. Evaluation was made based on the flanking sequence for each locus. After the preliminary evaluation with ADT file, Illumina provided Sequence list file where each locus was ranked with design-ability score and design rank (Figure 4).

The final rank of SNPs was made based on design-ability score (the rank of 0.6-1 indicating the high success rate) and the final rank for each SNP provided by Illumina. Additionally, flanking sequences for example with presence of insertions/deletions or with presence of degenerate nucleotides were removed as well as these with the length of sequence on either side of variant (central polymorphism) is less then 60bp. An annotation for each SNPs associated to the trait of interest/environment was also one of the priority criteria for the final rank of SNPs to be submitted for genotyping.

A total of 384 SNPs representing 242 genes were submitted to Illumina Genotyping Center in Parco Technico Padano, Lodi (Italy). The 384 SNPs were genotyped by using Illumina GoldenGate 384 SNP array where 188 SNPs originated from Arborea project, 94 SNPs originated from CRSP and 102 SNPs from Uppsala (Sweden) dataset (Supplement 2).

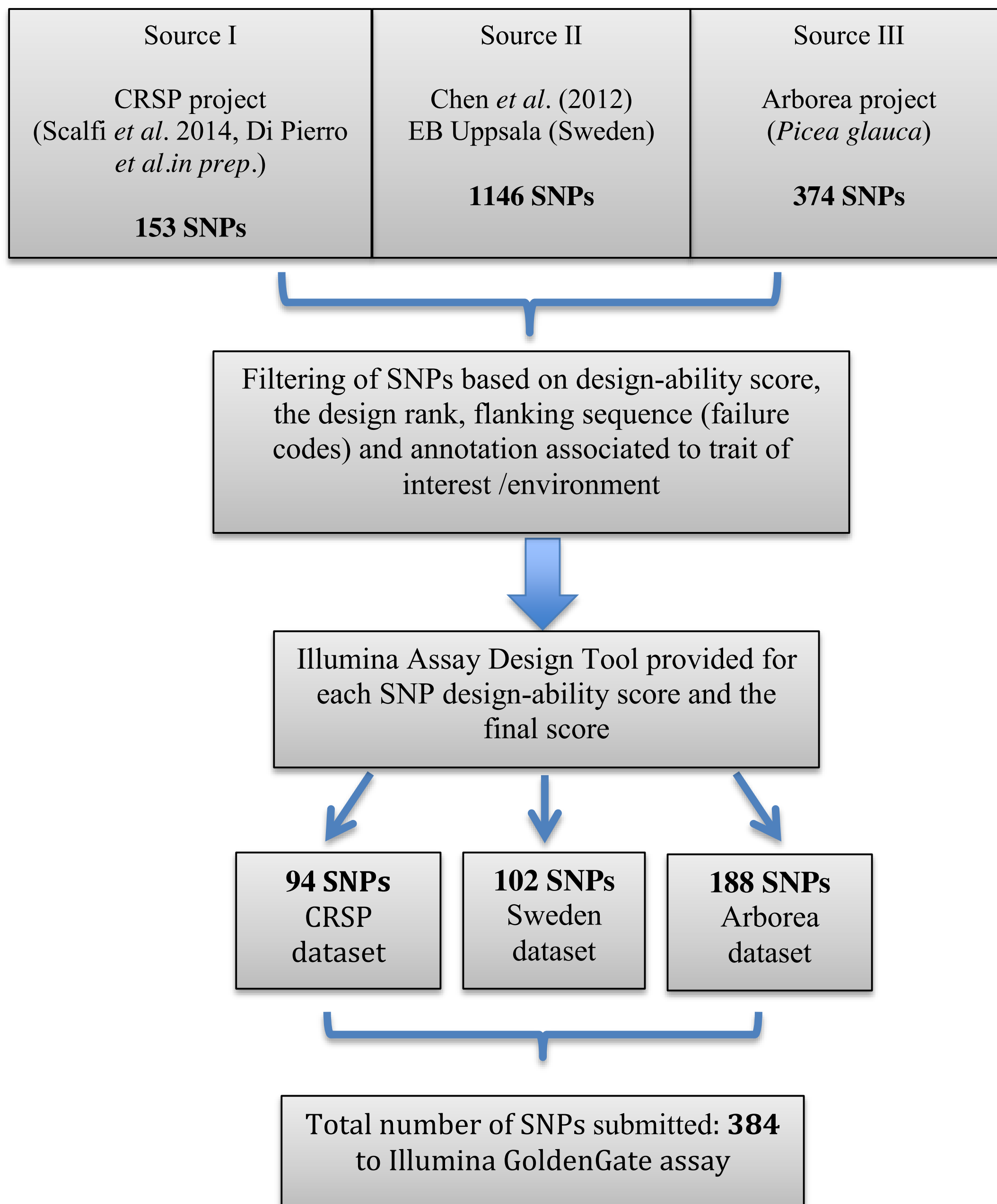



Figure 4: A 384 Illumina GoldenGate design steps

### 3.3 *Germination of seeds*

Seeds were placed in a sealable plastic bag and stored in the refrigerator at temperature 4°C shortly before starting the germination process. Norway spruce seeds were previously soaked in distilled water before placing them on the moisture filter paper inside Petri dish (Figure 5). For each individual were selected a group of 10-12 seeds.

Each Petri dish with seeds for the germination were treated with distilled water every 24h in order to keep filter paper regularly moisture (Figure 6). The germination lasted from 5-10 days resulting into slightly opened seeds and embryo development visible in only few mm. Seeds were collected in a single tube for each individual and used for extraction of the haploid tissue (megagametophytes). Total number of individuals used for germination of seeds was 309 (127 individuals from Austria and 182 individuals from Switzerland). Germination of Norway spruce seeds was done at Institute of Plant Genetics (IGV-CNR), Sesto Fiorentino (Italy) at Dr. Vendramin's laboratory.

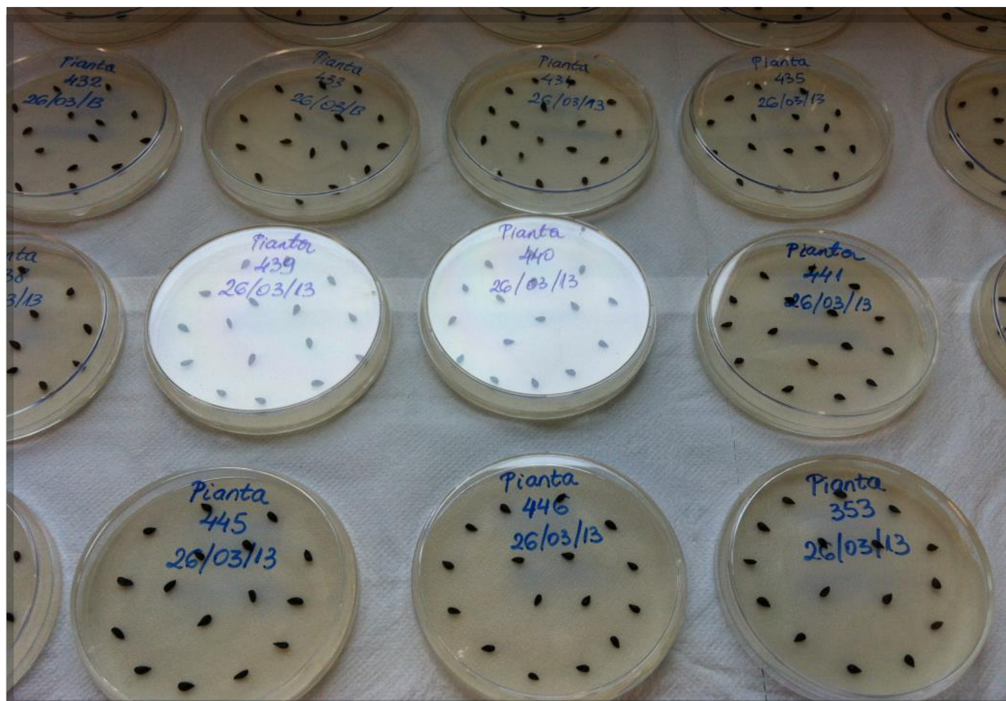



Figure 5: Germination of *Picea abies* seeds

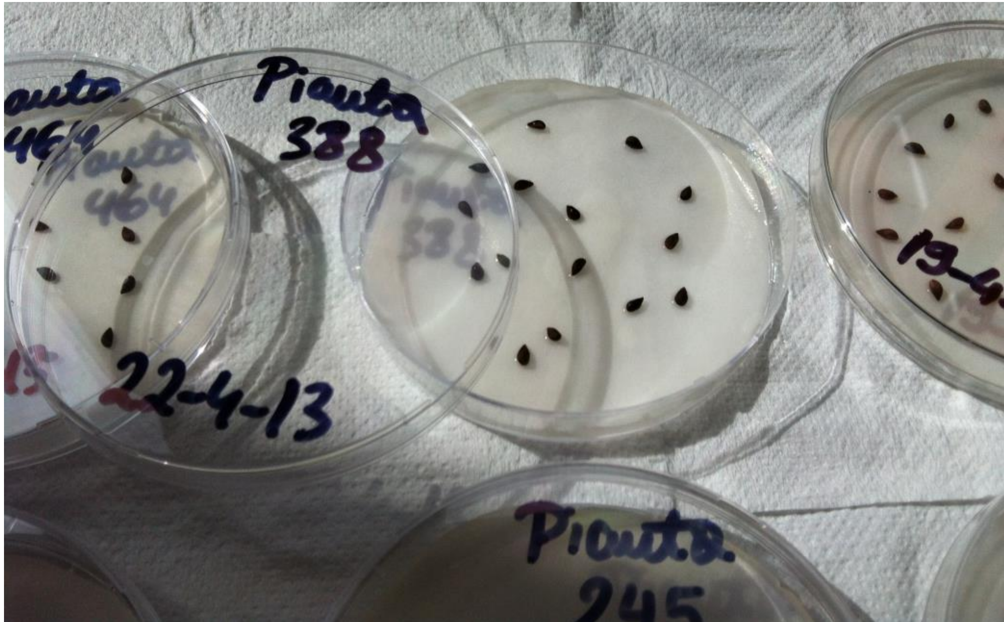



Figure 6: Germination of *Picea abies* seeds

The minimum number of pooled megagametophytes taken from the single cone was 8, and maximum was 10. As this procedure haploid tissue of spruce seeds is combined to diploid tissue in a single tube for each individual. Inferring the parent genotype from the combined haploid genotypes of mitotically derived products, it is possible that one of two alleles at a locus for which the parent is heterozygous is not detected if used only one megagametophyte from a single individual. However, with a set of 10-pooled megagametophytes per genet, the probability of such event is very small, taking into consideration that the allele of only one megagametophyte could be detected.

In some cases, the germination rate was quite low and minimum required number (8) of megagametophytes could not be achieved due to the fungal infestation or seeds were empty. The number of failed cases of the successful germination was 14 among individuals from Austria and 5 from Switzerland. Additionally, individuals that were germinated again were 52 from Switzerland and 12 from Austria due to the very low germination rate.

### 3.4 *DNA isolation from needles and megagametophytes*

Genomic DNA was extracted from both needles and megagametophytes (haploid tissue of *Picea abies* seeds) using the DNeasy plant Mini Kit (QIAGEN, Valencia, CA). Total number of individuals where needles were used for DNA extraction was 211 (individuals from Slovenia, Italy and France) whereas megagametophytes were used for 112 individuals (Austria) and 174 individuals (Switzerland).

Prior DNA isolation, plant material (needles) was weighted on the balance for each individual. Approximately 30mg of plant material was used per individual for DNA isolation. Regarding megagametophytes, at least 8 per individual were used for DNA isolation. Plant material was stored in

2ml micro-centrifuge tubes with 2.0 mm diameter beads overnight at -80°C. Beads, which were used for grinding of material, were previously washed and sterilized before use. Disruption of plant material was done at grinding machine until plant material did not become a fine tissue powder (approximately 1minute ~ needles and 30 seconds ~ megagametophytes). After disruption of plant material, it was added Buffer AP1 (400µl) and RNase (4µl) followed by cell lyses of the mixture for 1h at 65C with mixing (inverting samples) every 15minutes. When cell lyses were completed, a Buffer P3 (130µl) was added to mixture and incubated on ice for 5 minutes. The next step was centrifugation the lysate at 14,000-rpm providing removal of precipitates and cell debris. The volume of 450µl of lysate was recovered followed by direct addition of Buffer AW1 (675µl) on lysate and immediate mixture. From the mixture, 650µl was pipetted into DNeasy Mini spin column placed in 2ml centrifuge tube and centrifuged for 1min at 8000rpm. DNeasy Mini spin column was re-used into a new 2ml centrifuge tube when added Buffer AW2 (500µl) and centrifuged for 1min at 6000 rpm. This step was done one more time, however second time centrifuge was done on 14,000 rpm to dry a membrane. DNeasy Mini spin column was transferred to a new 2ml centrifuge tube when added 50µl of Buffer AE directly on membrane followed by incubation for 10minutes on the room temperature (15-25°C) and then centrifuged for 1 minute on 8000rpm to elute.

Loading DNA on agarose gel was done to estimate integrity for each sample and to test for a genomic DNA quality. Additionally, DNA concentration was measured for each sample with Qubit™ fluorometer (Invitrogen) in order to achieve minimum quantity of genomic DNA required by Illumina GoldenGate assay (20ng/µl).
DNA isolation procedure was done at Institute of Plant Genetics (IGV-CNR), Sesto Fiorentino (Italy) at Dr. Vendramin's laboratory.

## *3.5 Environmental dataset*

Environmental dataset was obtained for each sampled tree, provided by geo-reference data (longitude, latitude and altitude) by using two different sources. The first was WORLDCLIM dataset (Global Climate data), which provides monthly average data for the period from 1950-2000, with a spatial resolution 1km (Hijmans *et al*., 2005). Additionally, PGIS MODIS LST (Land surface temperature) satellite (http://modis.gsfc.nasa.gov), which provides daily data, was used with a spatial resolution 250m x 250m for the period 2002-2012 (Neteler, 2010) (Figure 7). Environmental data set was obtained in the collaboration with GIS and Remote Sensing Unit at FEM San Michele (Italy).

Monthly and daily environmental variables, temperature (°C) and precipitation (mm), were obtained for the period from 1980-2012 based of geographic variables for each site. Monthly minimum, maximum and averages were used to calculate seasonal measures for both temperature and precipitation. Seasonal measures were estimated based on annual quarters composed by tree months for each season as winter: (December-February), spring (March-May), summer (June-August) and autumn (September-November). Mean annual temperature (MAT) and mean annual precipitation (MAP) measures were obtained by averaging monthly measures for period 1980-2012 (Table 4).

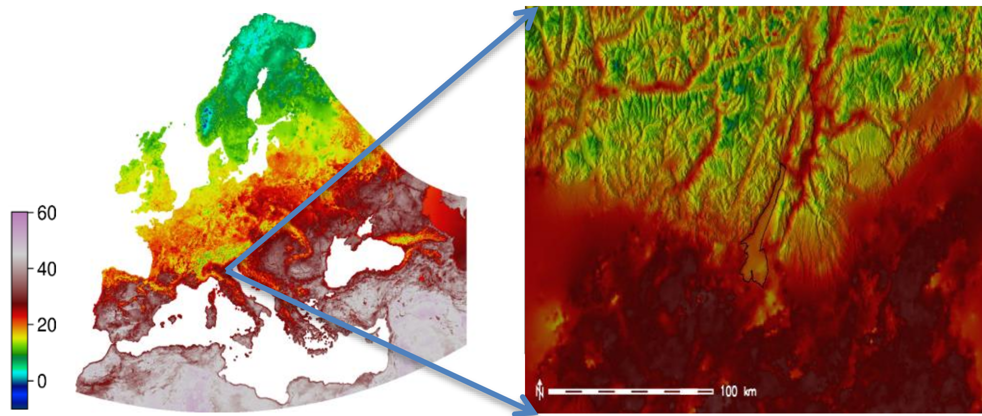



Figure 7: PGIS Modis LST satellite provides resolution of 250m x 250m (Neteler, 2010)

Table 4: List of environmental variables.

| Variable | Description | Measure Unit |
|---|---|---|
| Tmax seasonal | Seasonal maximum temperature | (°C) |
| Tmin seasonal | Seasonal minimum temperature | (°C) |
| Tmean seasonal | Seasonal average temperature | (°C) |
| Tmax monthly | Monthly maximum temperature | (°C) |
| Tmin monthly | Monthly minimum temperature | (°C) |
| Tmean monthly | Monthly average temperature | (°C) |
| PPT monthly | Monthly precipitation | (mm) |
| MAP | Mean annual precipitation | (mm) |
| MAT | Mean annual temperature | (mm) |
| AI monthly | Monthly values of aridity | |
| AI seasonal | Seasonal values of aridity | |
| Elevation | Elevation computed with GIS | |

### *3.5.1 Aridity index (AI) calculation*

Monthly temperature and precipitation data were used to estimate potential evapotranspiration (PET) based on the method of Thornthwaite (1948).

$$PET_i = \left[\frac{2\,u\,N}{45}\right]\left[\frac{10\,T_i}{I}\right]^{(6.75\mathrm{x}10^{-7})I^3 - (7.7\mathrm{x}10^{-5})I^2 + (1.79\mathrm{x}10^{-2})I + 0.49}$$

Where *u* is the average monthly day length in hours for month *i*, *N* is the number of days in month *I*, $T_i$ is the mean monthly temperature, and *I* is the heat index. The heat index is calculated as the sum across months of $(T_i/5)^{1.541}$. Monthly mean temperatures were used for the estimation of heat index.

The length of each day varies with location upon the Earth and day of the year. Day length definition is explained as the time between the beginning of sunrise, when the upper rim of the sun is apparently even the horizon until the end of sunset, when the upper rim of the sun is apparently even horizon (Harrison, 1960).

For the most accurate modeling of day length, several parameters have to be taken into account such as: position of the Earth with respect to the Sun, the rotation and orbital revolution of the Earth must be modeled and position of the flat surface on which the light is incident on the Earth (Forsythe *et al*. 1995).

An average day length (including twilight) from latitude (L), longitude (L) and the sun's declination angle for each day in the year was estimated by using modified Schoolfield's equations by using a day length coefficient of p=0.8333, providing the most accurate estimation of day length (Forsythe *et al*. 1995).

Formula derivation for day length at a point on the Earth at elevation zero with non-sloping ground can be divided into three main parts (Schoolfield, 1982):

1. Predicting the revolution angle (θ) from the day of the year (J)

$$\Theta = 0.2163108 + 2\tan^{-1}[0.9671396 \tan[0.00860\,(J-186)]] \qquad (1)$$

2. Predicting the sun's declination angle (Φ) or the angular distance at solar noon between the Sun and the equator, from the Earth orbit revolution angle

$$\Phi = \sin^{-1}[0.39795 \cos \Theta] \qquad (2)$$

3. Predicting day length (D) (plus twilight) from latitude (L), longitude (L) and the sun's declination angle.

$$D = 24 - \frac{24}{\pi} \cos^{-1} \frac{\sin \frac{p\,\pi}{180} + \sin \frac{L\,\pi}{180} \sin \Phi}{\cos \frac{L\,\pi}{180} \cos \Phi} \quad (3)$$

where p is in degrees and θ and Φ are in radians.

4. Predicting potential evapotranspiration (PET)

$$PET=(2\mu N / 45)\,(10T_i / I)^{(6.75\ 10^{-7})\,I^3 - (7.7\ 10^{-5})\,I^2 + (1.79\ 10^{-2})\,I + 0.49}$$

Where µ presents average monthly day length in hours for month i and I presents heat index, which is calculated as $I=(T_i / I)$.

Day length (*u*) once calculated is included into formula of evapotranspiration (PET). Aridity index (AI) was defined as the ratio of monthly precipitation to PET with this ratio being defined quarterly.

### *3.5.2 Principal component analyses*

Principal component analysis (PCA) is a tool to reduce multidimensional data to lower dimensions while retaining the most of the information (Kind and Jackson, 1999). The basic goal of principal components analysis is to describe variation in a set of correlated variables, $x^T = (x_1,\ldots x_q)$, in terms of a new set of uncorrelated variables, $y^T = (y_1,\ldots y_q)$, each of which is a linear combination of the x variables (Everitt and Hothorn, 2011). The final aim of principal component analysis is retaining as much as possible of the variation present in the data set.

The reduction of data is achieved by transforming to a new set of variables, the principal components, which are uncorrelated and which are ordered so that the first retain most of the variation present in all of the original variables (Jollite, 2002). PCA analyses produce linear combinations of the original variables to generate the axes, also known as principal components, or PC's (Holland, 2008).

Principal component analyses could be generally performed either with covariance matrix which scaled sums of squares and cross products or with correlation matrix when variables are previously standardized. Correlation matrix is always used when used different variables, which are measured in different units. Principal components should only be extracted from the sample covariance matrix when all the original variables have the same scale (Everitt and Hothorn, 2011).

A principal component analysis was applied using *prcomp* function in R software. The "*prcomp*" function is based on correlation matrix structure. The data were scaled and centered to avoid encountering of problems during performance of PCA analyses that might occur due to the significant difference of the relative magnitude of variables compared to the respective values. The reason is that linear regression favors those variables that show greatest variance, which would be usually those with larger values (reference or better write).

R function "*prcomp*" performs a principal component analysis using a singular value decomposition of the data matrix. The data matrix does not need to be centered or scaled prior the analyses because these options could be specified as arguments in the *prcomp* command by using "center" (a logical statement to indicate the data should be centered at zero) and "scale" (a logical statement to indicate the data should be scaled, i.e. the variance was set to 1 for each variable) (Anderson, online tutorial).

The final outputs important to be considered for further analyses are certainly PC's scores (also referred as loadings), the proportion of variance, cumulative variance explained and eigenvalues (Table 5).

Since the purpose of principal component analyses is the reduction of dimensionality (focusing on a few principal components versus many variables), the decision of how many scores should be selected and how many ignored, depended on several criteria. There are three main criteria to select the most important PC's. One of the most common one is to ignore principal components at the point at which the next PC offers little increase in the total percent variance explained. A second criterion is to include all those PC's, which explained total variance of 90%. The third criterion is to ignore components whose variance explained is less then 1 (when correlation matrix applied). The last criteria is to ignore the last PC's whose variance explained is all roughly equal (Holland, 2008).

The climate dataset offered totally 12 environmental variables measured and estimated for the period from 1980 until 2012. Monthly, annual and seasonal climatic data, longitude and latitude provided for each mother tree, were applied for PCA statistical analyses to summaries the major variation of the information contained into many dimensions into reduced number of uncorrelated dimensions. The number of the final PC's selected was based on the percentage of the variance they explained (e.g. at least 80%) and those for which eigenvalues (the variance explained by each component) were higher than 1, reducing the data from 20 variables (in case of seasonal data) to 3 variables (PC's). Monthly, seasonal and annual datasets and their combinations (monthly-seasonal; monthly-annual etc.) were also applied for PCA in order to obtain the highest possible variance.

Table 5: Outputs of "*prcomp"* function to do principal component analysis in R

| prcomp(x) | Interpretation |
|---|---|
| **sdev*sdev** | Eigenvalues |
| **rotation** | Eigenvectors |
| **x** | Scores |
| **sdev** | Standard deviations of each column of the rotated data |
| **center** | Mean value used for centering |
| **summary(x)$** | Proportion of variance explained |

### *3.6 Genotyping Data Analyses*

Genotyping of SNPs was performed by Illumina GoldenGate platform utilizing BeadArray™ platform for 384-plex assays at Parco Tecnologico Padano in Lodi, Italy. The process of genotyping was initialized by submission of final assay panel that were included designs predicted to have a high likelihood of success for genetic analysis experiments.

GoldenGate assay was described by Fan *et al*. (2003) explaining design steps. In particular, assay performs allelic discrimination directly on genomic DNA (gDNA), generates a synthetic allele-specific PCR template afterward, and then performs PCR on the artificial template. Genomic DNA (gDNA) is attached to a solid support prior the start of the assay. Assay oligonucleotides targeted to specific SNP (single nucleotide polymorphism) of interest are annealed to gDNA, since two allele specific oligonucleotides are designed for each SNP. Correctly hybridized oligonucleotides remain on the solid phase, whereas non-specifically hybridized oligonucleotides are removed by washing. For each SNP, two allele-specific oligonucleotides (ASOs) and one locus-specific oligonucleotide (LSO) are designed. After washing and annealing, an allele primer extension is carried out with DNA polymerase, followed by ligation of extended ASOs to their corresponding LSOs to create PCR templates. Lastly, primers that are fluorescently labeled, each with different dye, are added for two alleles for each SNP and generating a fluorescent image for each genotype (i.e. AA as homozygous, AT as heterozygous and TT and homozygous) (Fan *et al*. 2003).

All genotypes are visualized in GenomeStudio software v. 3.1.3.0 (Illlumina Inc. http://www.illumina.com). Each SNP has been analyzed independently to determine the genotype. Prior clustering of SNPs, preliminary quality sample evaluation has been made for samples, which required removal of failed samples. Additionally, GenomeStudio parameters have been optimized in order to achieve the highest genotyping accuracy.

First optimized parameter was gencall cutoff or no-call threshold, a quality metric calculated for each genotype, which ranges from 0-1. The gencall threshold of 0.25 was set up prior SNPs clustering meaning that genotype scores below this threshold were not assigned as genotypes since they were considered to be too far from the cluster and they were assigned a "no call" for that particular locus. Additionally, for each SNP has been optimized GenCall Score, a quality metrics, which indicates the reliability of the genotypes called. A GenCall score value is calculated for every genotype and it was set up to 0.45, and generally it could range from 0-1.

Each SNP was analyzed independently to identify genotypes. Manual edition for each SNP was done based on the angle of the clusters, dispersion of the clusters, overlap between clusters and their intensity, after what they have been classified as successfully and non successfully clustered SNPs. Secondly, they have been categorized based on the segregation rate on polymorphic and monomorphic ones. Minor allele frequency (MAF) is helpful to identify loci where homozygotes were incorrectly identified. It can range from 0-1 and all SNPs with Minor Frequency less then 0.1 should be evaluated as potentially false homozygotes, however the threshold of 0.5 was used as a minimum threshold to classify SNPs as monomorphic ones.

Clustering of each SNP was done by manual edition based on quality metrics and on criteria referred. It was carefully examined each choice of manual edition, always taking into consideration the The Law of Segregation ("The First Law").

### *3.7 Basic diversity statistics of genotype data*

SNP genotype data are provided into a sample x SNP matrix where samples are listed across columns and SNPs are listed down the rows provided by full table report from GenomeStudio software v2011.1 (Illumina Inc. http://www.illumina.com). This is particularly one of default formats to export data from GenomeStudio software. Within each cell into matrix, are presented diploid genotypes in the form allele1/allele2 (e.g. AT). The matrix of genotype data contained all 384 SNPs, which were included into genotyping as well as 400 individuals.

Firstly, this genotype data matrix was used to identify the number of successfully genotyped individuals and those failed during genotyping. Additionally, failed and successfully genotyped SNPs were identified. For every failed SNPs, was identified their source, annotation, final rank score and disagnability score.

Secondly, SNP data table report was extracted from GenomeStudio software v2011.1 (Illumina Inc.). The SNP Table column Minor Freq measures the SNP minor allele frequency and could help identify loci where homozygotes were incorrectly identified. Minor Freq values are ranging from 0-1. Based on minor allele frequency (MAF), which was obtained for every single SNP, classification was done on monomorphic, polymorphic and rare SNPs. All SNPs with minor allele frequency (MAF) less than 0.5

were discarded from further basic diversity statistics analyses, which were assumed not to be statistically important since their genotypes are considered to be monomorphic and not to provide genetic information.

SNP data quality was estimated based on call frequency (Call Freq), which presents the proportion of all samples at each locus with call scores above no-threshold and which range from 0-1. Additionally, $GC_{50}$ score (50%_GC_Score) was considered as indicator of SNP quality. It represents the $50^{th}$ percentile of the distribution of GenCall scores across all called genotypes. Lastly, missing data for each SNP were also considered as important indicator of SNP quality data also known as No_Calls presenting the total number of genotypes in each sample with a GenCall score below the no-call threshold as defined at the beginning of genotype data analyses.

Analyses of the basic diversity statistics of genotype data, sample x SNP matrix containing successfully genotyped individuals and polymorphic loci (loci with minor allele frequency ≥ 0.5) was done in R software v.3.0.2 (http://www.R-project.org). Prior the genetic analyses, it was launched a library package GENETICS v. 1.3.8.1 (Warnes *et al.* 2012), which provided an access to useful functions in population genetics. For each SNP was obtained allele frequency, genotype frequency, expected heterozygosity (He) and Poly. Inf. Content. However, obtained expected heterozygosity was additionally calculated by using a loop (developed by Andrew Eckert) and which provided output for major allele frequency, minor allele frequency, observed heterozygosity (i.e. the sample of heterozygote genotypes) and sample size for each SNP (i.e. number of genotyped samples minus missing data).

Further, calculation of sample Fis was based on Wright's definition of Fis = (He-Ho)/He, where He is expected heterozygosity and Ho presents observed heterozygosity. It indicates the average difference between observed and Hardy-Weinberg expected heterozygosity within each population due to the non-random mating. To calculate expected heterozygosity (He) was used function in R (developed by Andrew Eckert) and as result was obtained Fis across all SNP.

In the continuation, the same matrix data was used to calculate hierarchical fixation indices with R library package HIERFSTAT (Goudet, 2013), which allows estimation of hierarchical F statistics from haploid and diploid genetic data. F-statistics (hierarchical fixation indices) was performed in R software v.3.0.2 (http://www.R-project.org). Beside converted data, which were previously used in basic diversity statistics analyses, additional file with population identifiers was included into F-statistic analyses. Population identifiers were listed into one single file where all individuals were listed across the rows. Assignment of individuals into populations was done based on the assumption where each population represents a family (a group of individuals) at the same site (geo-referenced location). The total number of populations assigned was 152. Firstly, was calculated multilocus F-statistics, which was obtained in the matrix providing the total Fst for populations.

### *3.8 Population structure analyses*

A population consisted of 394 individuals successfully genotyped and 284 loci, was used to infer population structure and classify individuals into populations based on their genotypes. The main aim was to identify the number of populations where each population is characterized by allele frequencies at each locus.

Prichard *et al*. (2000) described a Bayesian clustering method for using multilocus genotype data to infer population structure and assign individuals (probabilistically) to populations. Bayesian method particularly accounts for the presence of Hardy-Weinberg within populations, complete linkage disequilibrium between loci within populations and attempts to find a population grouping that are not in disequilibrium (Pritchard *et al*. 2000). Since each population is modeled by allele frequencies, the knowledge about populations of origin of the individuals and allele frequencies in all populations is given by the posterior distribution (Pritchard *et al*. 2000).

A Bayesian model-based clustering method is implemented in the program STRUCTURE v.2.3.4 (http://pritchardlab.stanford.edu/structure.html) for inferring population structure using multilocus genotype data consisted of unlinked markers. It also provides identification of distinct genetic populations, assignment of individuals to populations and identification of migrants and admixed individuals. The model works well with the most common molecular markers such as SNPs (single nucleotide polymorphism), RAPD (random amplification of polymorphic DNA) or SSRs (microsatellites). This method and the software usage was introduced and described by Pritchard *et al*. (2010).

A final report from GenomeStudio software v.2011.1 (Illumina Inc.) was used to construct input file, which was launched into STRUCTURE v2.3.4 software. The format of the genotype data was arranged as a matrix in a single file, in which individuals were placed in the rows and the loci in columns. Each locus was stored in two columns for diploid organism, each one for one allele. Every genotype was coded by a numbers, where every nucleotide was numbered as following: Adenine (A)=1, Thymine (T)=2, Guanine (G)=3 and Cytosine (C)=4. Missing data (genotypes with no call) were numbered as a -9 (a number that does not occur elsewhere in dataset). File was saved in txt (text) format and launched in STRUCTURE software v.2.3.4.

The beginning of genetic clustering of individuals into populations was previously arranged by admixture model as ancestry model of preference with allele frequency correlation. Admixture model indicates that each individual has inherited some fraction of the genome from ancestors in population k reporting as posterior mean estimates of these proportions (Pritchard *et al*. 2010). Admixture model was useful to estimate population structure and explain the spatial repartition of genetic markers. Among parameters used to perform admixture model was the length of burnin period was selected to 100,000 with number of MCMC (Markov Chain Monte Carlo) Reps after burnin was set up to 500,000. Initial value of ALPHA (Dirichlet Parameter for Degree of Admixture) was set up to 1. Number of genetic clusters (K) from 1 to 10 was tested and for each value, it was run 10 independent MCMC simulations (10 iterations).

## *3.9 Outlier detection methods*

### *3.9.1 BayeScan outlier detection method*

The program, BayeScan is an approach, which provides identification of candidate loci under natural selection from genetic data using differences in allele frequencies between populations. It estimates the probability that each locus is under selection using a Bayesian method, which implements the multinomial-Dirichlet likelihood. BayeScan approach was described and developed by Foll and Gaggiotti (2008). The software BayeScan is available at http://cmpg.unibe.ch/software/BayeScan/download.html.

BayeScan method is based on an island model (Wright 1931) in which subpopulation allele frequencies are correlated through a common migrant gene pool from which they differ in different strength (Foll and Gaggiotti, 2008). The difference in allele frequency between common gene pool and each subpopulation is measured by a subpopulation Fst. In particular, this method estimates the posterior probability of given locus being under the selection, suggesting if positive that locus *i* is under directional selection, whereas if negative, locus *i* is under balancing selection. The evidence of selection occurring is provided by Reversible Jump Markov Chain Monte Carlo (RJ-MCMC) algorithm (Green 1995) that estimates the posterior probability (Foll and Gaggiotti, 2008).

BayeScan is command line software and it has been coded using standard C++ (UNIX platform). Input file for BayeScan was based on genotype matrix file after clustering of genotypes from Illumina GenomeStudio v.2011.1 (Illumina Inc. http://www.illumina.com). Genotyping matrix composed of 392 individuals and 280 loci, was used to create input file. Genotypes are labeled as AA (dominant), AB (heterozygous) and BB (recessive). Prior submission to BayeScan program, all genotypes was coded as numbers that provided exact input file when dominant genotypes (AA) were coded as “0” (zero), heterozygous (AB) were coded as “1” and recessive (BB) as “2”. Missing data or no calls was coded as “-9”.

For each population, total number of genes for particular locus was estimated based on number of individuals per population. The total number of genes represented twice the number of individuals for diploid data, considering also missing data. The number of observations for each of two alleles in each population was estimated. For each individual at each locus, the observed genotypes were coded as “0”, ”1” and “2”. This number summed the total number of genes for that particular population.

Since BayeScan is a population-based approach, all individuals were assigned in the populations. Each population was defined as a group of individuals sampled at the same geographic location (site) provided by geo-reference data (i.e. latitude, longitude) (Supplement 3). BayeScan was running under following parameters: Burn in: 5000, Thining interval: 10, Number of MCMC (Markov Chain Monte Carlo iterations): 1000000, Number of pilot runs: 20 and Length of each pilot run: 5000.

### *3.9.2 Bayesian linear mixed model (BayEnv*)

As outlier detection method, a Bayesian liner mixed model (Bayenv) has been applied to detect loci potentially involved in the local adaptation based on mapping population of Norway spruce. This model has been designed and described by Coop *et al*. (2010). Bayesian method estimates the empirical pattern of covariance in allele frequencies between populations from a set of markers and then uses this as a null model for a test at individual SNP (Coop *et al*. 2010). Bayenv software is available at http://www.eve.ucdavis.edu/gmcoop/.

The main aim was to perform a test of selection on genetic markers, in such case SNPs (single nucleotide polymorphism) based on identification of loci, which might show allele frequency correlation within one or more environmental variable. These loci could be under selection driven by environmental factors or correlation selection pressures (Coop *et al*. 2010).

Bayesian model was applied to the genotype data for 392 individuals sampled across Alps. It was applied an approach when all individuals were assigned in the populations based on their geo-reference data (e.g. latitude, longitude). All individuals were prior assigned in populations for each provenance, Italy-Slovenia-France (8 adjusted groups); Austria (5 adjusted groups) and Switzerland (7 adjusted groups). For pursuing Bayenv analyses, this time genotyping data were organized that individuals were assigned in 20 populations and all 280 loci were included. After estimation of covariance-variance matrix (for a million MCMC (Markov Chain Monte Carlo iterations), an average matrix was used for the second step when was estimated the correlation between environmental variables (included in ENVIRONFILE) with SNPFILE (contained of 280loci).

In the genotype data were included 280 loci, which were successfully genotyped. The program has been coded using standard C++ and designed to run in UNIX environment and it is basically divided in two main steps.

In the first step, the program estimated covariance-variance matrix. The covariance-variance matrix provides information how much allele frequencies co-vary among populations due to their shared ancestry or gene flow. Its role is basically to indicate how well populations are genetically divergent or estimating neutral population structure. Covariance-variance estimation was done based on genotype matrix, which contained allele counts across populations for each SNP. Allele counts were placed in the two columns, for each allele. The sum of allele counts was equal to the sample size for each SNP for each population. Covariance matrix was estimated for million iterations providing an output every 5000 iterations.

In the second step, the program was running separately on each SNP providing quick estimation of Bayes factors for the environmental variable of interest. The main aim was to investigate whether some loci are indicating allele frequencies having strongly correlation with environmental variables taking into account previously estimated neutral population structure (covariance-variance matrix).

Among required input files for pursuing the second step, SNPFILE was submitted, which presented count data for every allele for single SNP, being in the same order as they appeared in covariance

matrix. In SNPFILE were included all 280 loci. Secondly, ENVIRONFILE, is the file of environmental variables previously standardized (i.e. the mean was subtracted from the value of each variable, followed by calculation of difference between individual's score and the mean and then divided through by the standard deviation of the variable across populations). In ENVIRONFILE were included four principal components, geographic data (latitude, longitude and elevation) as well as all monthly environmental variables (monthly minimum temperature, monthly maximum temperature, monthly mean temperature, monthly average precipitation, and monthly aridity). The last input file was MATRIXFILE, which presented an average covariance matrix over million iterations. Bayes factors were estimated by using a million MCMC (Markov Chain Monte Carlo) iterations.

Although Bayenv was done with approach when individuals were assigned into 20 populations based on their provided geo-reference data (i.e. latitude, longitude and elevation), additionally it was applied individual based approach where each population represent an individual sampled at the geographic site (location) provided with geo-reference data (i.e. latitude, longitude, elevation).

Bayenv analyses were performed in UNIX environment and both steps were done by following command lines provided in the manual. The manual is available:
(http://www.eve.ucdavis.edu/gmcoop/Software/Bayenv/bayenv_manual.pdf).

### 3.9.3 *Spatial analyses detection (Samβada)*

Samβada software presents integrated software for landscape genomics analyses of large datasets. It was applied in order to access whether some markers are potentially associated with an environmental variable. Samβada is command line software, a successive version of MatSAM software developed several years ago by Joost *et al*. (2007).

The key features are the study of local adaptation correlated with environment and the measure of spatial autocorrelation in environmental and molecular datasets (Stucki and Joost, 2014). Samβada uses logistic regression to estimate the probability of presence of an allelic variant for a polymorphic marker given the environmental conditions of the sampling (Joost *et al*., 2007).

Samβada software intends to correlate allele frequency with environmental variables to look for signatures for selection following logistic model of regression. It uses logistic regression model to estimate probability of an allelic variant for a polymorphic marker for environmental conditions for a sampling locations. To fit the models, maximum likelihood approach is used. In this case was applied univariate model where the probability of presence is same at each location and is equal to frequency of genotype. Significance is assessed with both log-likelihood G ratio and Wald test.

The likelihood ratio or G statistics is

$$G = -2 \ln L / L'$$

, where L is the likelihood of the initial model (with a constant only) and L' is the likelihood of the new model including the examined variable.

The Wald statistics is

$$W = \frac{\hat{\beta} i}{\sigma (\hat{\beta} i)}$$

Where β is the maximum likelihood for parameter *i*; βi is the maximum likelihood estimate of the parameter βi and σ(βi) is an estimate of its standard error (Joost *et al.*, 2007).

For Samβada analyses were prepared input files for molecular and environmental data. Molecular data were consisted of genotyping data for 392 individuals and 280 loci, which were extracted from GenomeStudio Illumina v.2011.1 (Illumina Inc. http://www.illumina.com). Each line in the file provided information of genotypes for each individual. Molecular data set used for analyses were arranged in the form of matrices; each row of the matrix corresponded to a sampled individual, while the columns were organized according to the sampled individual's geographic coordinates and contained binary information (1 or 0) depending on presence or absence of allele at locus in question. For each SNP was recorded presence of absence of alleles when each allele if present for a given individual was set as "1" and if absent allele, was set as "0". The results were tree columns for each SNP presenting all tree genotypes (i.e. AA as dominant was recorded as "1" "0", AB as heterozygous was recorded as "1" "1" and BB as recessive was recorded as "0" "1"). Since there were totally 280 loci successfully genotyped, the total number of columns in molecular data set was 840 (280*3=840). Molecular markers were all screened for minor allele frequency and filtered only those where maf >0.5 considering these as polymorphic ones. The number of SNPs after pruning (filtering) was 208 loci. For each locus were recorded tree versions of genotypes and totally were assigned 624 columns in the molecular data set.

Environmental file contained a dataset for seasonal variables. Totally 21 environmental variables including geo-reference data (latitude and longitude) for each sampled individual, were included into analyses. Among seasonal environmental variables, a seasonal minimum temperature; a seasonal maximum temperature; a seasonal mean temperature; a seasonal precipitation and seasonal aridity were included for the period from 1980 until 2012. Environmental file was pruned (filtered) to keep all variables showing low correlations. Variables showing high correlations with others might lead to over-significant threshold with the Bonferroni correction. For instance, "t_max_win" (temperature maximum winter) showed to be highly correlated with other temperature variables and if included all temperature variables, models might result in very similar results.

Additionally, a parameter file was prepared, which is needed as well to set up the analyses. The parameter file contained one line per parameter where specified following: the number of environmental variables was 25 (where 21 columns were environmental variables, 2 columns presented geo-reference data such as longitude and latitude, and two other columns presented the state where sampled and id number for each individual); the number of molecular markers was 625 (where (208*3)

plus one column presenting id for sampled individual); the number of individuals included in the analyses (392). The next parameter was option which indicates which column contain identifiers of individuals when id code was set up as "id" followed by subset of environmental variables defined with only those showing high correlation and further processed "ppt_sum", "ppt_aut", "t_max_win", "altitude", "Ai_spring". Further parameter was a model chosen for processing analyses as dimmax=1 indicating a model with a constant and 1 explanatory variable and as a final one parameter was selected best model indicating that only save significant univariate models and rank them according to Wald score. The threshold applied was $2.67*10^{-6}$ (= 0.01/(208*3*6)) where 0.01 presents a threshold, 208 presents all loci with maf ≥ 0.5, tree different scenarios for each SNP were presented with 3 and number of environmental variables only considered to be evaluated in the model was 6.

Samβada software is available http://lasig.epfl.ch/sambada. It is a command line program designed in C++ (UNIX environment). Program is launched with an indication of the correct path to Samβada folder and listing of all input files along previously stored in the same folder.

# 4. RESULTS

## 4.1 *Climatic data estimation within Norway spruce population*

Principal component analyses (PCA) were performed on the subsets of environmental variables in order to estimate the percentage of proportion of variance. Each subset was independently used to reduce its dimensionality as well as combination of subsets. PCA analyses were performed in R software with "*prcomp*" function.

Among all environmental subsets, all three main ones (i.e. seasonal, monthly and annual) had a proportion of variance higher than 90% as well as their combinations. The main aim was to estimate which one would have the highest proportion of variance indicating the most variation extracted from the dataset. This parameter was important one in decision which dataset to include in outlier detection analyses.

As the final choice to be included, seasonal environmental dataset was selected since it's proportion of variance was higher than 90% and it presents the simplest dataset. It is composed of thirty different variables grouped in five different categories ("ppt_seas"- seasonal precipitation; "AI_seas"- seasonal aridity index; "tmax_seas"- maximum seasonal temperature; "tmean_seas" – mean temperature seasonal; "tmin_seasonal" – minimal seasonal temperature).

In general, no strong correlation existed between climate data and geographic variables (e.g. latitude, longitude and elevation) and among climate variables (Figure 8). Longitude was strongly and negatively correlated to precipitation variables (Pearson's $r < -0.72$) and aridity index variables (aridity index for winter and aridity index for summer, Pearson's $r < -0.67$), whereas low and positively with temperature variables (Pearson's $r < 0.47$). Latitude was also negatively correlated to precipitation variables (Pearson's $r < -0.38$) and aridity index variables (aridity winter and aridity summer, Pearson's $r < -0.41$) and low and positively to temperature variables (Pearson's $r < 0.43$). Elevation was strongly and negatively correlated to temperature variables (Pearson's $r < -0.91$) whereas moderately and positively with precipitation (Pearson's $r < 0.56$).

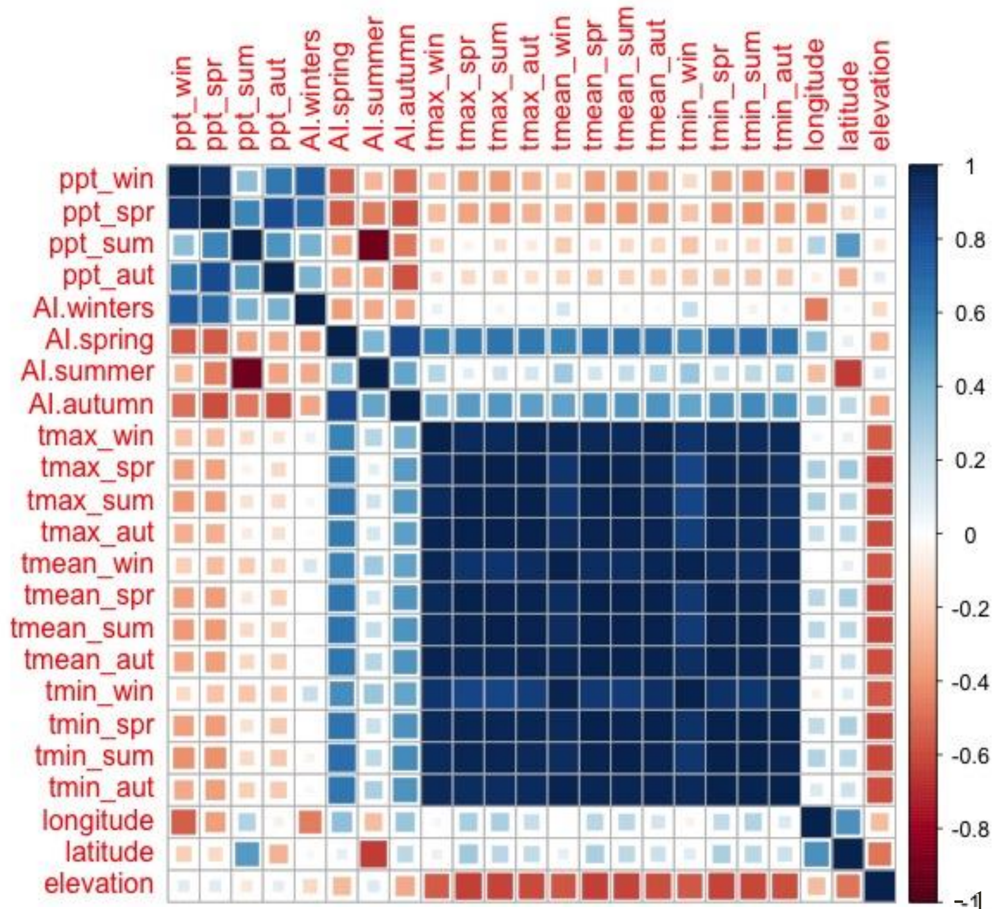



Figure 8: Visualization of correlation matrix between 20 climatic variables and geographic variables reveals no strong and generally negative correlation among variables. Variables abbreviations were listed on left axis and on the top. The correlation plot was generated in R studio software v. 098.51 by using library "*corrplot*".

Among environmental variables, the correlation was generally positive and moderate. Particularly, the correlation between aridity index and temperature variables was from moderate to strong and positive (Pearson's *r* from 0.47 up to 0.90) whereas exception was aridity for winter, which was negatively correlated (Pearson's *r*….). Temperature and precipitation were negatively and strongly correlated as well (Pearson's *r* < - 0.79) whereas temperature and aridity index for all seasons were moderately to strongly correlated (Pearson's *r* from 0.50 - 0.90). Aridity index and precipitation were strongly and negatively correlated (Pearson's *r* < - 0.98) where an exception was aridity index for winter to precipitation, which were positively and strongly correlated (Pearson's r < 0.88).

The top three principal components (PCs) captured the most (i.e. 90%) variance for the climatic variables included in dataset (Table 6). These top three PCs also had eigenvalues greater than 1 (Figure 9) and cumulative proportion of variance was indicated for each (Figure 10). For every PCs, loadings were obtained explaining an arrangement of variables for each PCs (Table 7). In particular, for the first PCs aridity index as well as temperature variables were loaded, for the second only aridity index for

spring, summer and autumn, whereas for the third PCs precipitation for winter, spring and autumn; aridity index for summer, autumn and winter; mean temperature for autumn and minimum temperature for autumn and winter. The cumulative proportion of variance was ranged from 64% for the first PCs, 83% for the second PCs and 90% for the third PCs (Figure 9).

Table 6: Summary statistics of the PCA on climatic variables presenting the importance of principal components explaining the distribution for the proportion of variance and cumulative proportion if variance for top three principal components.

| Parameter | PC1 | PC2 | PC3 |
|---|---|---|---|
| **Standard deviation** | 3.58 | 1.95 | 1.18 |
| **Proportion of variance** | 0.64 | 0.19 | 0.06 |
| **Cumulative proportion** | 0.64 | 0.83 | 0.90 |

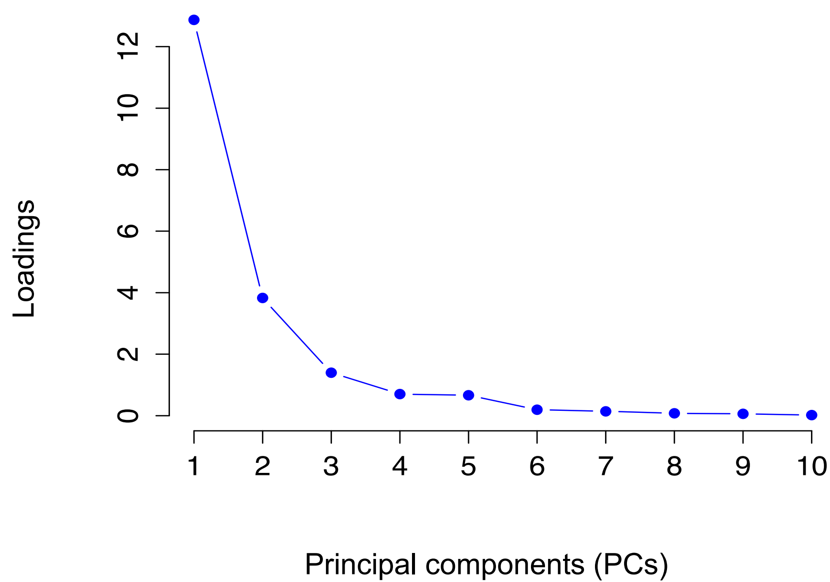



Figure 9: Visualization of loadings for each principal component indicating the first three are having loadings greater than 1.

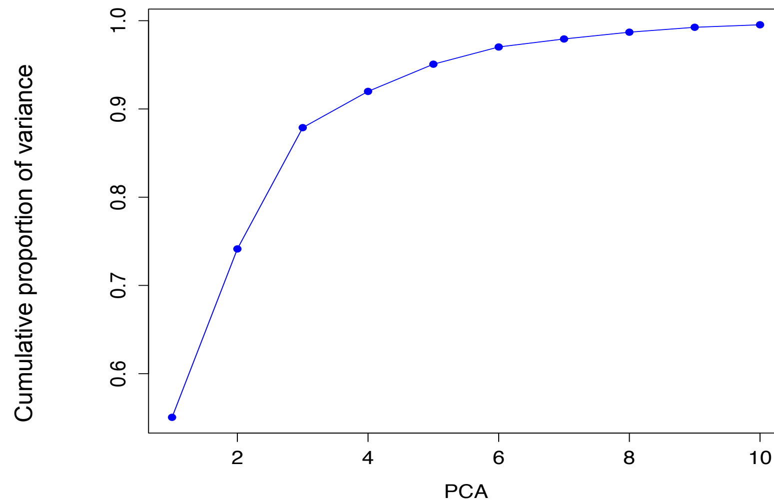



Figure 10: Visualization of the cumulative proportion of variance (%) for each principal component subsequently increasing for the first three PCs.

Table 7: Multivariate measures of climate for the seasonal dataset

| PC* | Eigenvalue | PVE⁺ | Description |
|---|---|---|---|
| 1 | 12.87 | 64 | Spring, summer, autumn aridity, temperature |
| 2 | 3.83 | 19 | Spring, summer and autumn aridity |
| 3 | 1.39 | 6 | Winter, spring, autumn precipitation<br>Winter, summer, autumn aridity<br>Winter maximum temperature<br>Autumn mean temperature<br>Winter and autumn minimum temperature |

* Principal component.

⁺ Percent variance explained.

Additionally, Principal component analyses (PCA) were performed on the monthly subset of environmental variables in order to estimate the percentage of proportion of variance. The monthly subset was independently used to reduce its dimensionality. PCA analyses were performed in R software with "*prcomp*" function. The monthly dataset contained totally 60 variables grouped as "ppt" (precipitation)/year, "ai" (aridity)/year, "tmax" (maximum temperature)/year, "tave" (average temperature), "tmin" (minimum temperature)/year.

The top four principal components (PCs) captured the most (i.e. 90%) variance for the climatic variables included in dataset (Table 8). These top four PCs also had eigenvalues greater than 1 (Figure 10) and cumulative proportion of variance was indicated for each (Figure 11). For every PCs, loadings were obtained explaining an arrangement of variables for each PCs (Table 9).

In particular, for the first PCs aridity index as well as temperature variables were loaded, for the second only aridity index for months from March until October, whereas for the third PCs precipitation for June, July, August and September; aridity index for months from April and May; maximum temperature for all months except for November and December, average temperature for months from March to September, minimum temperature for months from April to September. For the fourth PCs, precipitation for months January, February, March, July and August, aridity index for January, February, September, November and December, average temperature for January, April, May, October, December and minimum temperature for all months except for June, July and August.

The cumulative proportion of variance was ranged from 62.1% for the first PCs, 81.4% for the second PCs, for the third PCs 89% and 92% for the fourth PCs (Figure 9).

Table 8: Summary statistics of the PCA on climatic variables presenting the importance of principal components explaining the distribution for the proportion of variance and cumulative proportion if variance for top three principal components.

| Parameter | PC1 | PC2 | PC3 | PC4 |
|---|---|---|---|---|
| **Standard deviation** | 6.10 | 3.40 | 2.14 | 1.45 |
| **Proportion of variance** | 0.62 | 0.19 | 0.07 | 0.03 |
| **Cumulative proportion** | 0.62 | 0.81 | 0.89 | 0.92 |

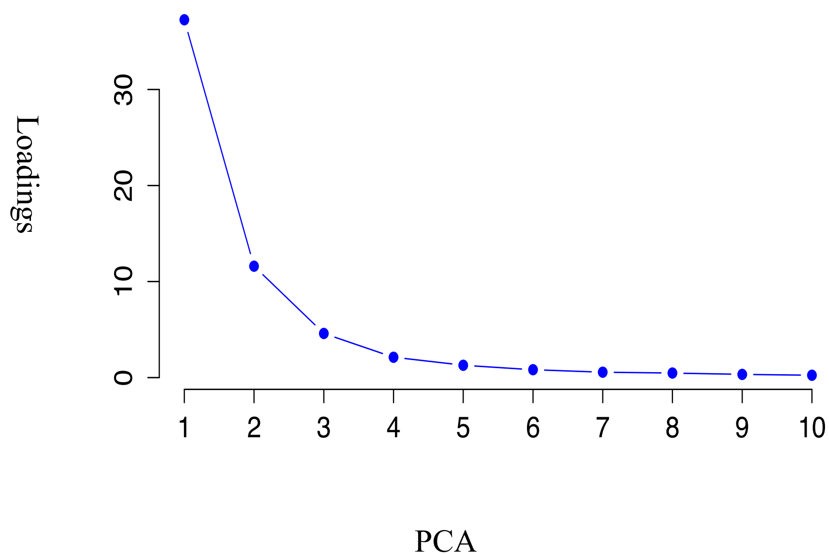



Figure 11: Visualization of loadings for each principal component indicating the first four are having loadings greater than 1.

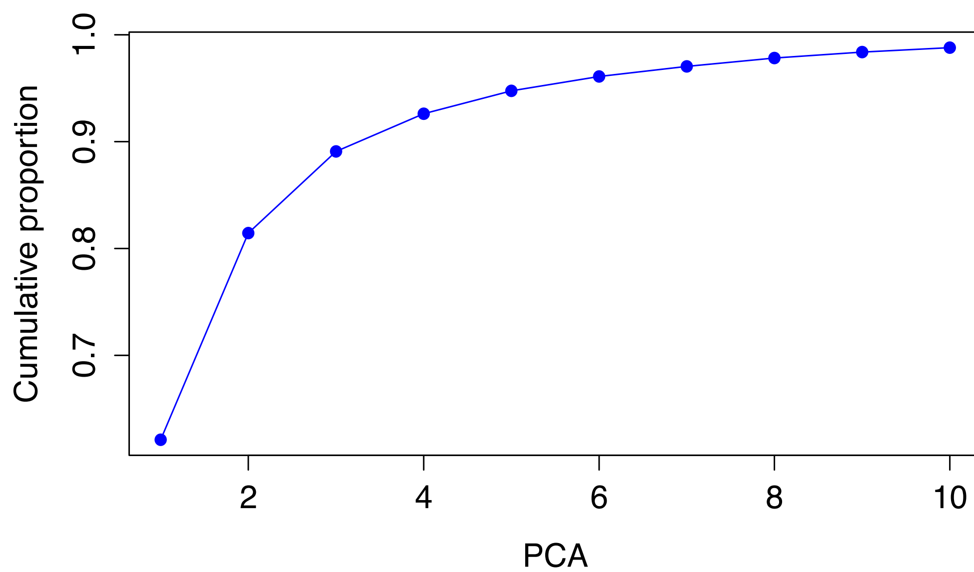



Figure 12: Visualization of the cumulative proportion of variance (%) for each principal component subsequently increasing for the first four PCs.

Table 9: Multivariate measures of climate for monthly dataset

| PC* | Eigenvalue | PVE+ | Description |
|---|---|---|---|
| 1 | 3.72 | 62 | Aridity, maximum temperature, average temperature, minimum temperature for all months |
| 2 | 1.16 | 19 | Aridity for March-October |
| 3 | 4.59 | 7 | Precipitation from June-September, Aridity for April and May, Maximum temperature for February-October, Average temperature for March to September, Minimum temperature for April to September |
| 4 | 2.11 | 3 | Precipitation for January to March and July to August, Aridity for January to February, September, November, December, Average temperature January, April, May, October and December, Minimum temperature for all months except June, July and August |

## 4. 2 *Analyses of 384 Illumina GoldenGate genotyping data*

Analyses of genotyping data were performed in GenomeStudio Illumina v.2011.1 (Illumina Inc. http://www.illumina.com) software when each SNP was analyzed independently to determine genotypes. Total number of genotyped individuals were 400 and successfully genotyped 394 whereas 6 individuals failed due to the low quality of genomic DNA. From total 384 SNPs genotyped, clusters were reliable and genotypes were successfully assigned in case of 284 SNPs (Figure 13).

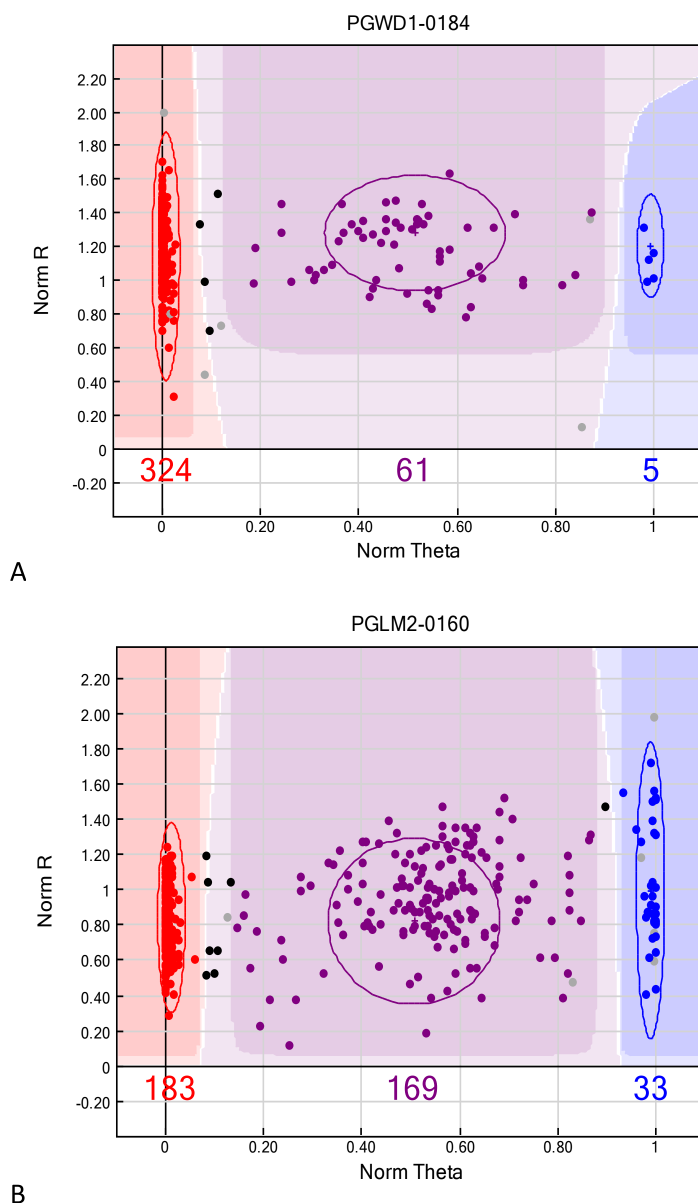



Figure 13: Examples of successfully clustered SNP providing a clear separation between three clusters when genotypes are well distinguished (A, B)

The number of SNPs failed during genotyping was 100. In case of these SNPs, clustering was ambiguous due to the formation of diffuse/poor cluster formation with unreliable genotypes and such SNPs have been zeroed (Figure 14). Additionally, SNPs were also zeroed when their frequency was too low to be attributed to a potential biological effect.

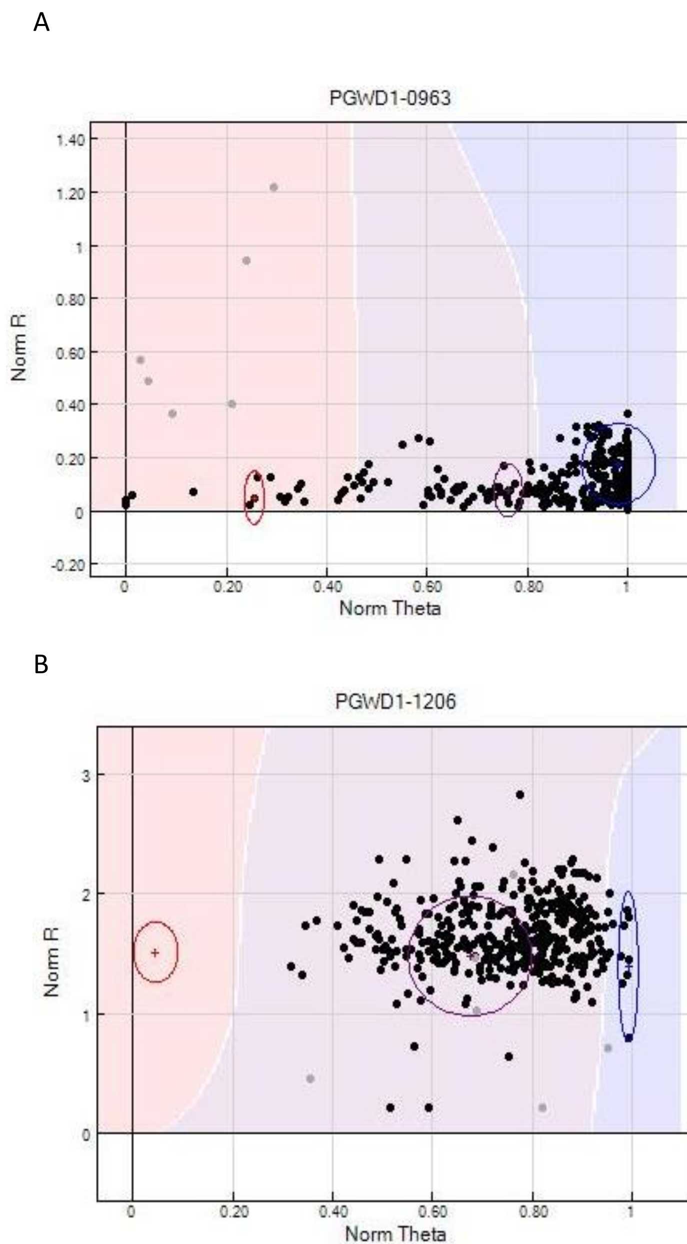



Figure 14: Examples of unsuccessfully clustered SNPs showing low call frequency (A) and overlapping of clusters with unreliable genotypes (B).

Additionally, remaining SNPs were classified based on minor allele frequency on polymorphic (208), monomorphic (47) and rare (29) SNPs. In particular, as polymorphic SNPs were assigned when maf (minor allele frequency) ≥ 0.05; as monomorphic SNPs when maf < 0.05 and as rare SNPs when maf < 0.01 (Table 10). Minor allele frequencies measures were extracted from the SNP table provided by GenomeStudio Illumina v.2011.1 (Illumina Inc. http://www.illumina.com).

Table 10: Basic diversity statistic of genotyping data from 384 Illumina dataset

| 384 Illumina chip | Number | Percentage |
|---|---|---|
| SNPs failed during genotyping | 100 | 26% |
| Polymorphic SNPs* | 208 | 54% |
| Monomorphic SNPs* | 47 | 12% |
| Rare SNPs* | 29 | 7% |

* Estimation based on minor allele frequency for each SNP

Regarding SNPs that failed during genotyping, the highest number (63) originated from Arborea project (Canada). The Arborea project was conducted on specie *Picea glauca* (White spruce) that is common for boreal forest of Canada and Alaska. Same SNPs were applied in study of Chen *et al.* (2012), when SNPs obtained from Arborea project were tested in 12 diploid individuals of Norway spruce from 12 populations in central Europe (Latvia, Poland and Germany). In particular, SNPs that came from Arborea project proved to be unsuccessful when genotyping on *Picea abies* due to non-existence of SNPs in a mapping population of Norway spruce (Table 11).

Additionally, 27 SNPs from Uppsala (Sweden) specie *Picea abies* (Norway spruce) and 10 SNPs from CRSP re-sequencing panel were also not successful during genotyping. Design-ability score was equal to one, predicting successful performance in genetic analyses (Supplement 4).

Table 11: Overview of unsuccessfully genotyped SNPs

| SNPs source | Species | Number | Final score (average) | Design-ability score (average) |
|---|---|---|---|---|
| Canada | *Picea glauca* | 63 | 0.824 | 1 |
| CRSP reseq. | *Picea abies* | 10 | 0.864 | 1 |
| Uppsala (SW) | *Picea abies* | 27 | 0.83 | 1 |

Overall the quality for SNP genotype data was overall high (e.g. the median Call Freq is equal to 0.97) indicating that the proportion of all samples at each locus with call scores above the no-threshold. In addition, the mean value for $GC_{50}$ score (50%_GC_score) was 0.70 indicating that there was a high percentage of distribution of GenCall score across all called genotypes (Table 12). The missing data per SNP (the median) was 8.30.

Table 12: SNP genotype quality data

| **SNP genotype quality data** | **Median** |
|---|---|
| Call frequency | 0.97 |
| Missing data per SNP | 8.30 |
| $GC_{50}$ score | 0.70 |

### *4.3 Basic diversity statistics on genotype data*

The calculation of basic diversity statistics was done based on genotyping matrix composed of only polymorphic SNPs (maf ≥ 0.5). It has been generated following summaries based on entire genotyping matrix: MAF (major allele frequency), which average, was equal to 0.76; maf (minor allele frequency), which average, was equal to 0.23; He (expected heterozygosity), which average was equal to 0.3262; Ho (observed heterozygosity), which average was equal to 0.3258 (Table 13). Both expected and observed heterozygosity were used to calculate Wright's Fis inbreeding coefficient across all loci, which formula is defined as Fis=(He-Ho)/He (Figure 15). The mean value of Fis estimated based on 208 polymorphic loci was equal to 0.0004 (Supplement 5).

Table 13: Basic diversity statistics of genotype data

| **Basic diversity statistics** | **Mean** |
|---|---|
| MAF – major allele frequency | 0.76 |
| maf – minor allele frequency | 0.23 |
| Hexp – expected heterozygosity | 0.3262 |
| Hobs – observed heterozygosity | 0.3258 |
| Fis across all loci | 0.0004 |

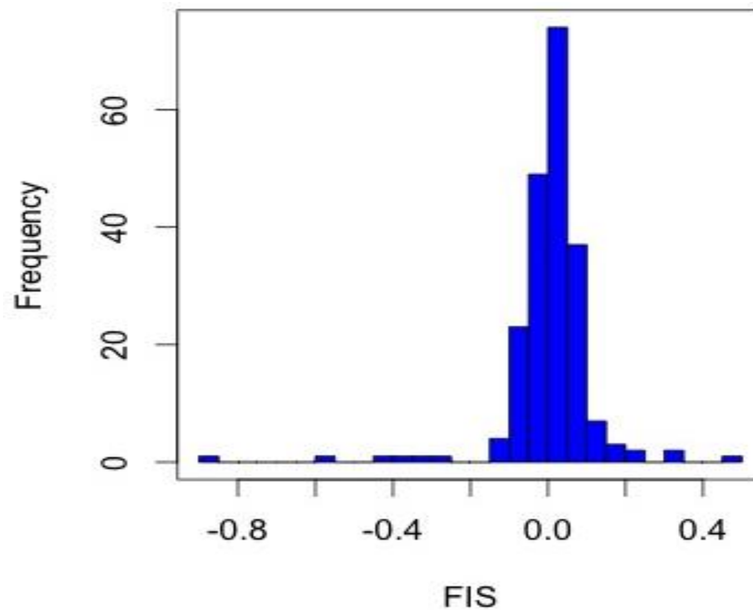



Figure 15: Visualization of Wright's Fis inbreeding coefficient across all loci where Fis=(He-Ho)/He

Fis compares average observed heterozygosity of individuals in each subpopulation and the average Hardy-Weinberg expected heterozygosity for all subpopulations (the "i" stands for individuals and "s" stands for subpopulations) thereby accounting for deviations from Hardy-Weinberg expected genotype frequencies due to the two main reasons such as excess or deficit in heterozygotes due to the non random mating within populations either due to the possible deficit of heterozygotes among subpopulations compared to panmixia. Fis value (fixation index per sample) for each SNP is the value to be considered before launching SNPs for the next analyses, having in mind of exclusion of all SNPs whose value is out of its optimal range (between -0.25 and +0.25). SNPs which were out of the optimal range counted for 59 SNPs having Fis <-0.25 and 78 SNPs having Fis > +0.25 leaving only 71 SNPs whose Fis values were in the optimal range. The number of SNPs for which Fis was within the optimal range, was quite low to retain, therefore the final decision was to include all SNPs into further analyses.

Fst represents a reduction in heterozygosity due to subpopulation divergence in allele frequency. It is equal to difference between averaged expected heterozygosity of subpopulations and the expected heterozygosity of the total population reflecting different degrees of allele frequency divergence among the sets of subpopulations. Fst range of values indicated less heterozygosity on average for subpopulations compared to heterozygosity expected to ideal case when the entire population is panmicitic (Figure 16A). The range of Fst values are ranging from -0.03 to 0.03. In addition, allele frequencies of subpopulations are slightly different and each has expected heterozygosity less than ½. In such case when heterozygosity of total population equal to 0.5 (maximum) that means there is no allele frequency divergence between two subpopulations (Figure 16B).

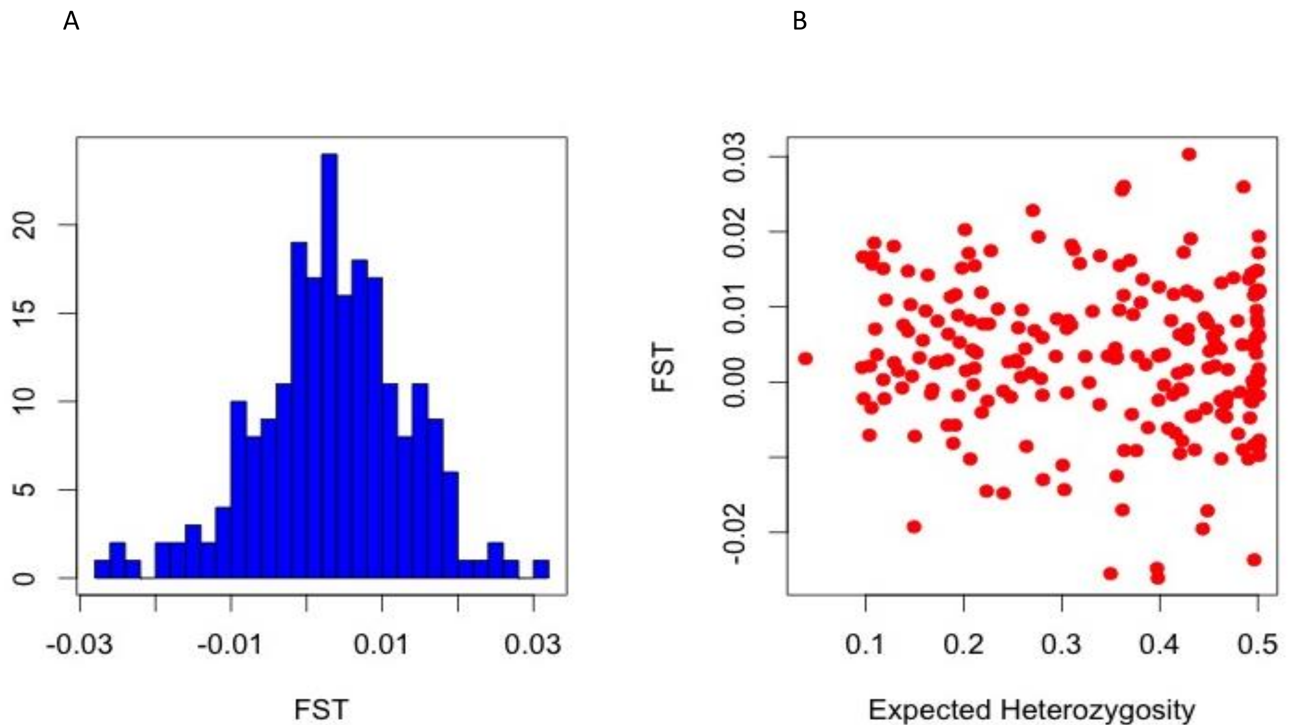



Figure 16. (A) Fst range of values indicating no presence of genetic variation within populations. (B) Expected heterozygosity range of values are less than 0.5 indicating no divergence in allele frequency between subpopulations.

## 4.4 *Population structure estimation*

An attempt to identify the number of clusters was based on a mapping population consisted of 394 individuals and 284 successfully genotyped loci. The simulation summary of admixture model (the length of burnin period was selected to 100,000 with number of MCMC (Markov Chain Monte Carlo) Reps after burnin was set up to 500,000) was launched in STRUCTURE HARVESTER (available at http://taylor0.biology.ucla.edu/structureHarvester/), a web-based program for visualization and generating the results from software STRUCTURE (Pritchard *et al*. 2000). The program was developed and described by Earl and vonHoldt (2011).

STRUCTURE HARVESTER (Earl and vonHoldt, 2011) provided visualization of plot of the mean likelihood L(K) per *K* value including standard deviation bars to display likelihood variance based on output files from STRUCTURE (Pritchard *et al*. 2000) (Figure 16). Use of L(K) is one of two approaches to determine the best K. This approach is known as Wilcoxon test (Nonparametric test) and it is based on assumption when K is approaching a true value, L(K) becomes plateaus (or continues increasing slightly) and has a high variance between runs. In this particular case, L(K) did not show clear mode of true K.

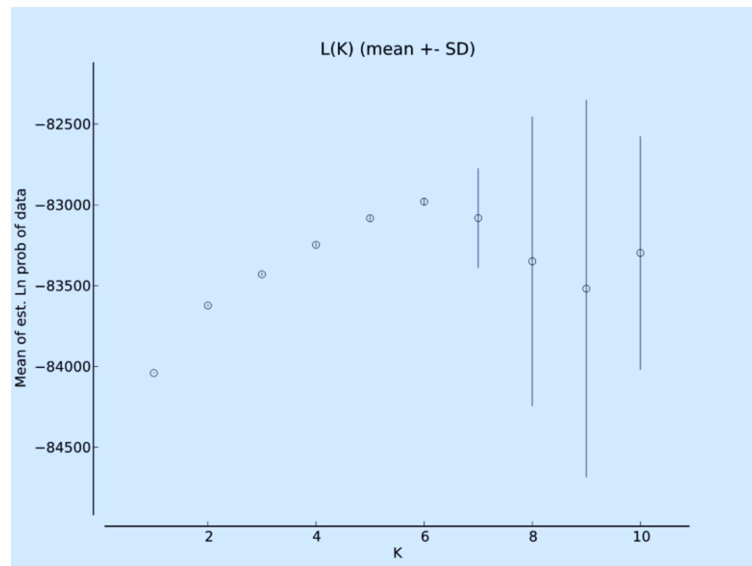
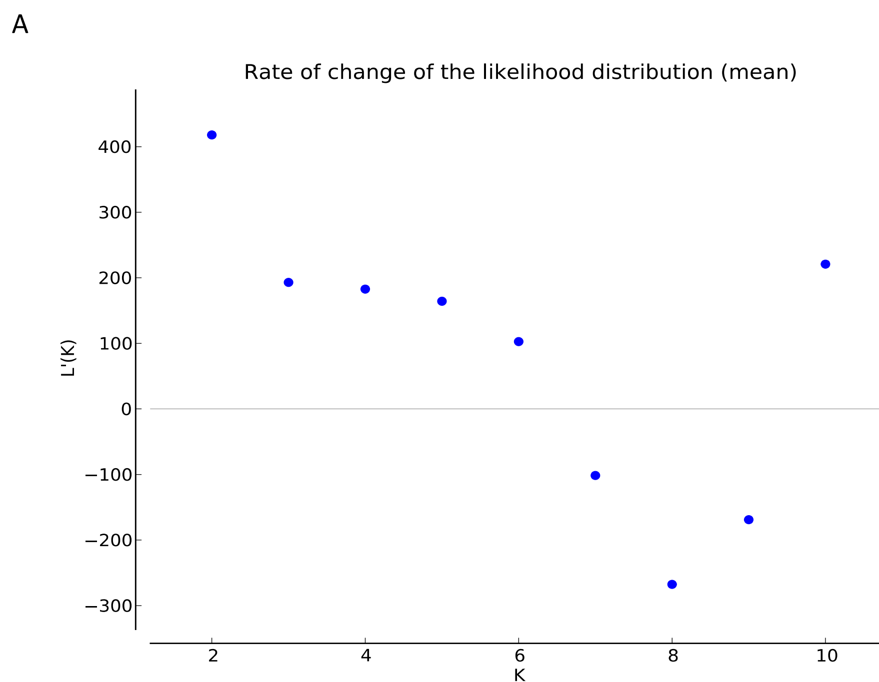



Figure 16. Plot of mean likelihood L(K) and variance per K value from STRUCTURE on dataset consisted of 394 individuals genotyped for 284 loci.

Additionally were provided a plot of rate of change of the likelihood distribution (mean ±SD) calculated as L'(K)=L(K)-L(K-1) and plot of the absolute values for the second order rate of change of the likelihood distribution (mean±SD) calculated according to the formula [L''(K)]=[L'(K+1)-L'(K)] (Figure 17).

A



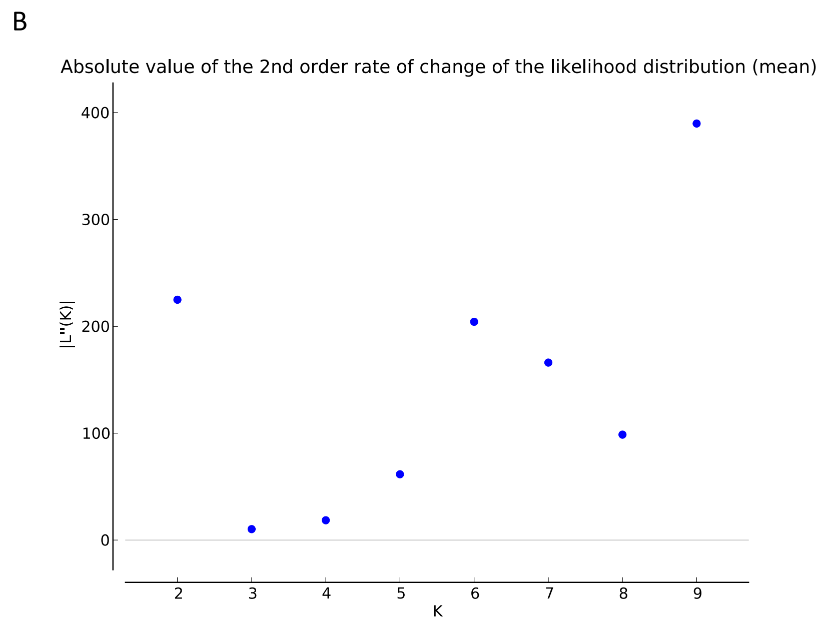



Figure 17. (A) The rate of change of the likelihood distribution (mean±SD) and (B) absolute values of the second order change of the likelihood distribution (mean±SD).

The use of an ad hoc quantity based on the second order of the rate of change of the likelihood function with respect to K (ΔK) is another method of estimating the optimal K, also known as Evanno method (Evanno *et al*. 2005). The Evanno method is defined as formula $\Delta K = m([L''K])/s[L(K)]$ where ΔK is estimated as the mean of the absolute values of L”(K) averaged over 10 runs divided by standard deviation of L(K). The ΔK shows a clear peak at the true value of K (Figure 18). The highest peak of ΔK was when K=2 indicating the presence of two main genetic clusters. The second highest peak of ΔK value was when K=6.

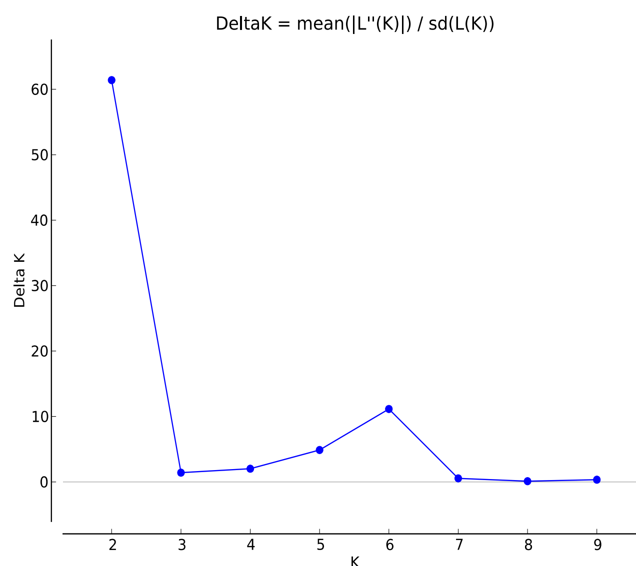



Figure 18. Plot of ΔK (Evanno method) for detection of the real number of K groups, which best fit, the dataset

Since Evanno method indicated presence of two main genetic clusters, the conclusion is that within a mapping population of 394 individuals is no strong population structure. The second highest rate of ΔK was when K=6 (Figure 18), however the membership of individuals in such clustering occasion, did not indicate any six cluster grouping, assuming there is an existence of panmixia or a population where process such as mating and individuals movement are uniform (Supplement 6).

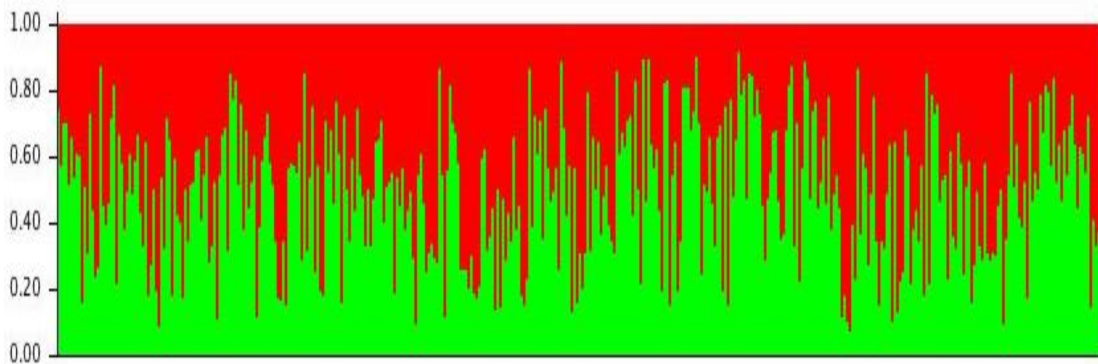



Figure 19. Barplot when K =2 for original order for 394 individuals assuming grouping two main populations. Each individual is represented by a single vertical line, which is partitioned into K colored segments that represent that individual's estimated membership fraction in each of K inferred clusters.

Table 14. Table output of the Evanno method results. Yellow highlight is indicating the largest value of the Delta *K* column. Data as in Figure 18.

| K | Reps | MeanLnP(K) | Stdev LnP(K) | Ln'(K) | \|Ln''(K)\| | Delta K |
|---|---|---|---|---|---|---|
| **1** | 10 | -84041.17 | 0.133749 | — | — | — |
| **2** | 10 | -83623.28 | 3.662968 | 417.89 | 224.89 | 61.395571 |
| **3** | 10 | -83430.28 | 7.24259 | 193 | 10.22 | 1.411097 |
| **4** | 10 | -83247.5 | 9.196255 | 182.78 | 18.51 | 2.012776 |
| **5** | 10 | -83083.23 | 12.623439 | 164.27 | 61.54 | 4.875058 |
| **6** | 10 | -82980.5 | 18.326726 | 102.73 | 204.25 | 11.144926 |
| **7** | 10 | -83082.02 | 305.105045 | -101.52 | 166.06 | 0.544272 |
| **8** | 10 | -83349.6 | 892.291293 | -267.58 | 98.65 | 0.110558 |
| **9** | 10 | -83518.53 | 1164.593194 | -168.93 | 389.78 | 0.334692 |
| **10** | 10 | -83297.68 | 719.453359 | 220.85 | — | — |

## 4. 5 *BayeScan outlier detection method*

BayeScan calculation ended with output file “prefixname_fst.txt”. In this file, each line corresponded to one locus and contained following values: (1) the index of the locus corresponding to the index in the input file; (2) the posterior probability for the model including selection; (3) the logarithm of Posterior Odds to base 10 for the model including selection; (4) the estimated alpha coefficient indicating the strength and direction of selection; (5) the Fst coefficient averaged over populations.

An R function provided to plot and identify outliers using output file “prefixname_fst.txt” when was calculated PO (Posterior Odds) threshold leading to a False Discovery Rate of no more than 5%. The outcome is the plot with outliers and their list. However, the plot merged did not indicate any outliers and the list was empty meaning that no outlier loci were detected within FDR of 5%.

As provided, Fst coefficient was obtained where in each population Fst was calculated as the posterior mean using model averaging (Supplement 7). Fst was plotted together with the logarithm of Posterior Odds to base 10 (which is fixed to 1000 when posterior probability is equal to 1). Fst averaged over populations indicated the highest value for 0.045 for the locus number 27 indicating its presence under purifying or balancing selection as it showed low level of genetic differentiation. The highest logarithm of the Posterior Odds to base 10 was equal to 0.76 for the locus PGWD-1034 (Arborea project, Canada) (Figure 20). The locus PGWD1-1034 is belonging to the AUX/IAA transcription factor gene family (auxin response regulator). The logarithm of the Posterior Odds corresponded to a Bayes factor equal to 3 indicating a “substantial” evidence for selection (Table 13). Posterior odds could be included in the Jeffrey’s scale of evidence for Bayes factors (Jeffrey, 1935, 1961). Posterior odds are defined as a ratio of posterior probabilities and indicate how likely the model with selection is compared to the neutral model (Foll, 2010).

$$PO = P(M2|N)/P(M1|N) = BF * P(M2)/P(M1)$$

, where M1 presents a model under selection, M2 presents a neutral model, N represents a given dataset, BF presents Bayes factor and P posterior probability.

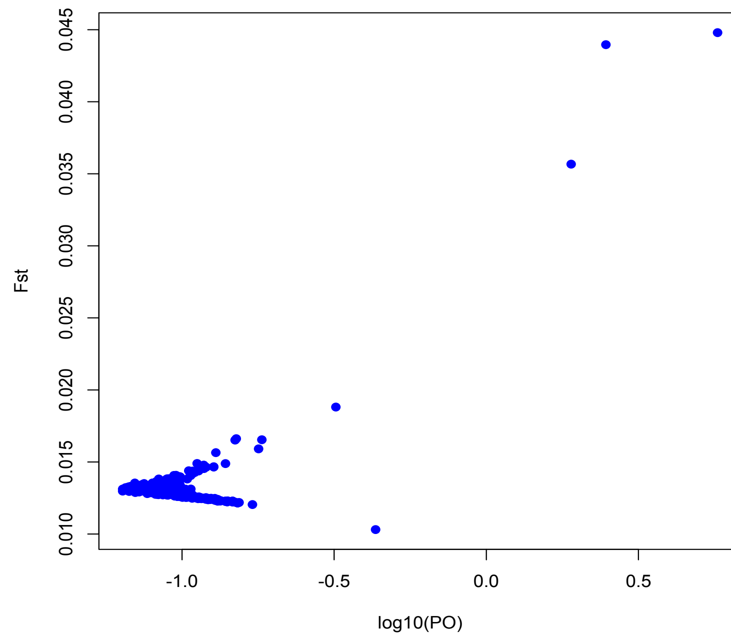



Figure 20. R plot visualization of logarithm for the Posterior Odds to base 10 and averaged Fst across populations.

Table 15. Interpretation of Jeffrey's scale of evidence for the selection (Jeffrey, 1961)

| Posterior probability | Bayes factor (BF) | log10(BF) | Evidence of the selection |
|---|---|---|---|
| 0.50 -> 0.76 | 1 -> 3 | 0 -> 0.5 | Barely worth mentioning |
| 0.76 -> 0.91 | 3 -> 10 | 0.5 -> 1 | Substantial |
| 0.91 -> 0.97 | 10-> 32 | 1 -> 1.5 | Strong |
| 0.97 -> 0.99 | 32 -> 100 | 1.5 -> 2 | Very strong |
| 0.99 -> 1.00 | 100 -> ∞ | 2 -> ∞ | Decisive |

For most of loci, the logarithm of the posterior odds was negative corresponding to non-existence of outlier loci potentially under the selection. In conclusion, BayeScan did not detect any outlier loci at 5% significance level given for 384 loci screened (Figure 21).

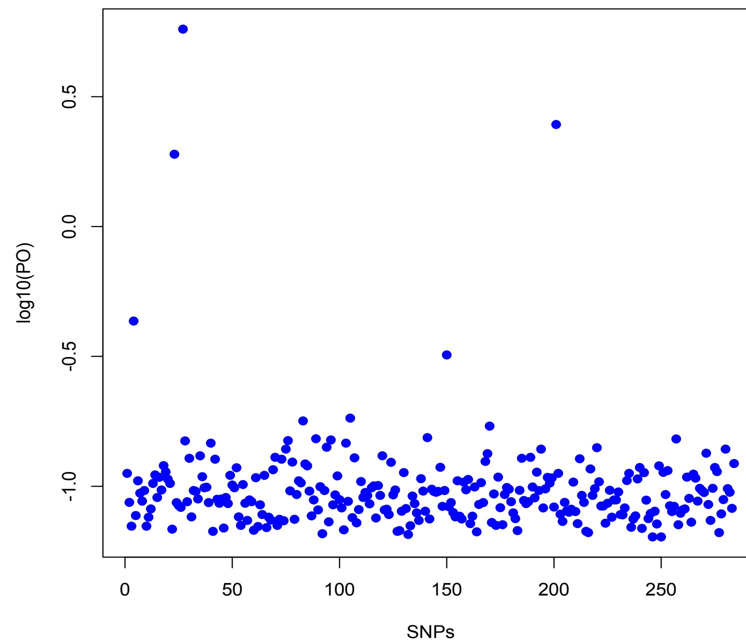



Figure 21. R plot visualization of logarithm for the Posterior Odds to 384 loci under the investigation for the signal of selection.

## 4. 6 *Environmental association analyses*: *Bayesian linear mixed model detection (BayEnv)*

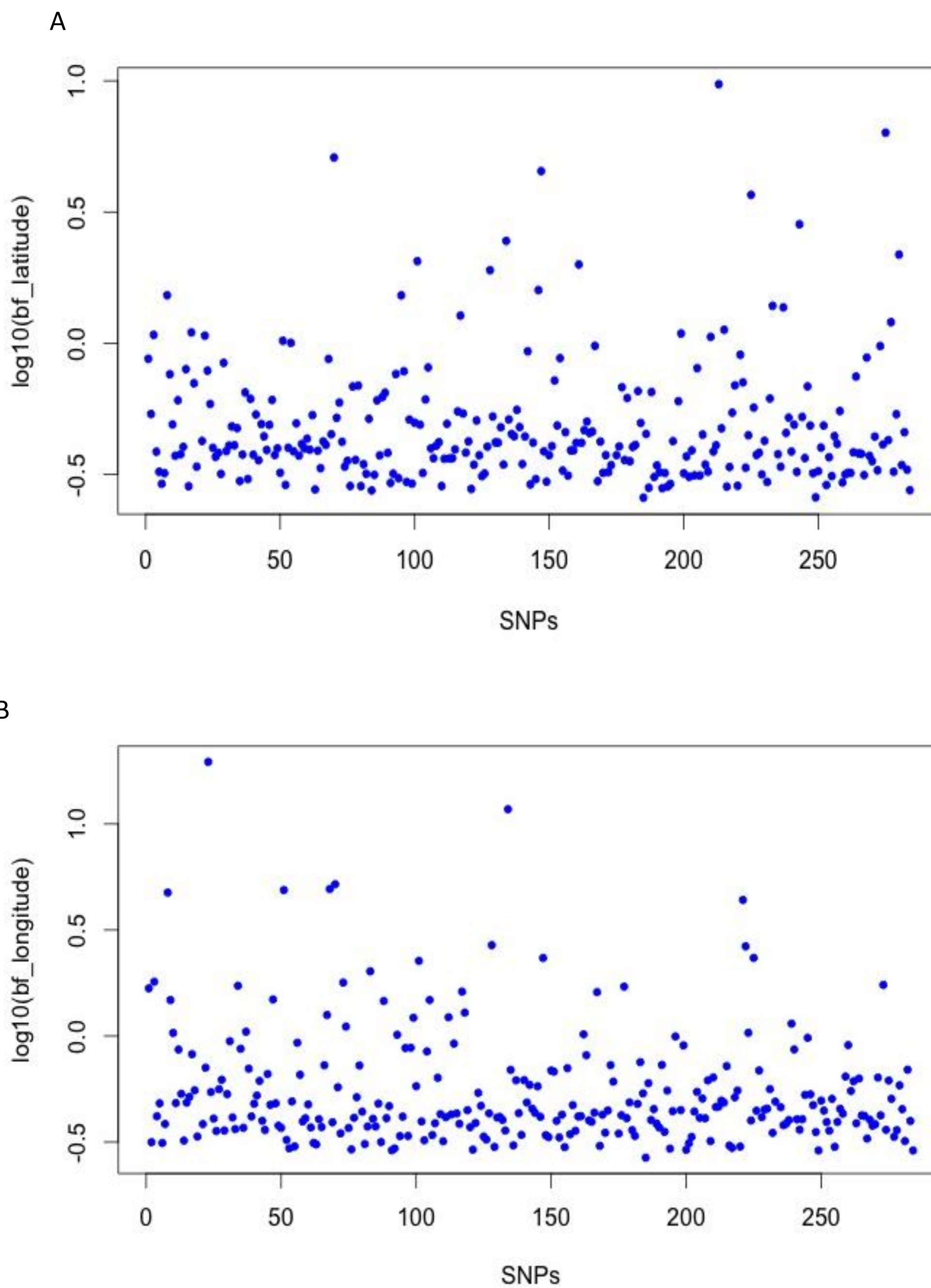
The first outcome file from BayEnv analyses was the covariance-variance matrix estimated over 20 populations and for million iterations. The rows and columns for each population were in the same order as has appeared in allele count file. The covariance-variance matrix was estimated each 5000 iterations until the final million iterations. For the further analyses, the mean covariance matrix over million iterations was used as input for the second step-environmental correlations analyses.

The covariance-variance matrix indicated how well populations are genetically differentiated due to the neutral process such as genetic drift and gene flow. Under the neutral model, the population allele frequency in each population may deviate away from an ancestral (or global) allele frequency due to the genetic drift. Some populations could be more genetically close to each other due to the effect of shared population history or gene flow. In such case, the neutral model is specified by covariance structure of allele frequencies across the populations (Coop *et al*, 2010).

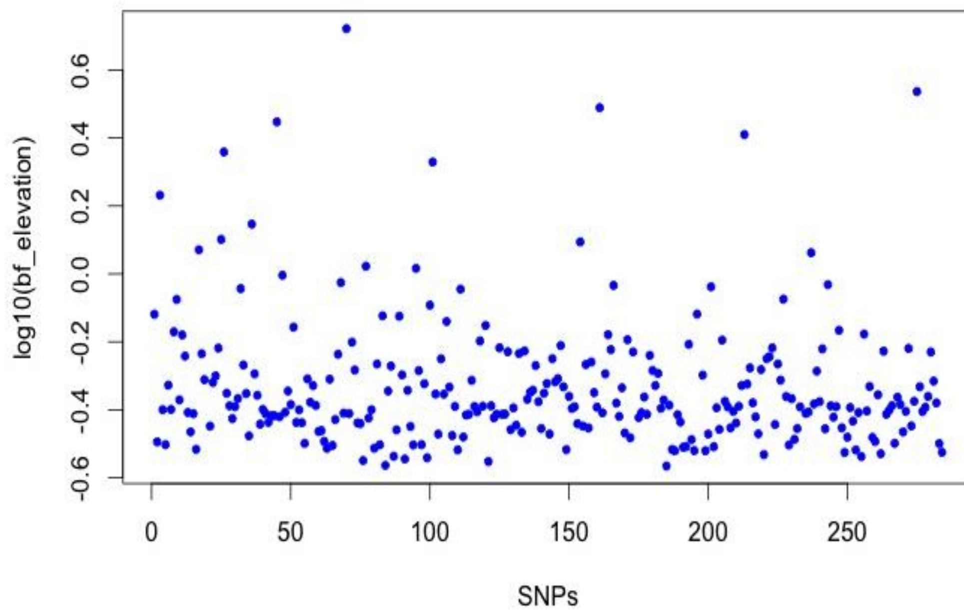
Next, it was explored the correlation between allele frequencies and environmental variables based on average covariance-variance matrix estimated in the first step. In this step, Bayes factors have been estimated for every SNP for each environmental variable across 20 populations and for a million iterations. An output file contained Bayes factors, which presented the measure of correlation strength between allele frequency and environmental variable. A Bayes factor (BF) presented a ratio of the posterior probability under the alternative model and the neutral (null) model. Every Bayes factor was

adjusted to the logarithm of base 10 for each SNP and was used for an interpretation within Jeffrey's scale of signal of the selection (Jeffrey, 1961).

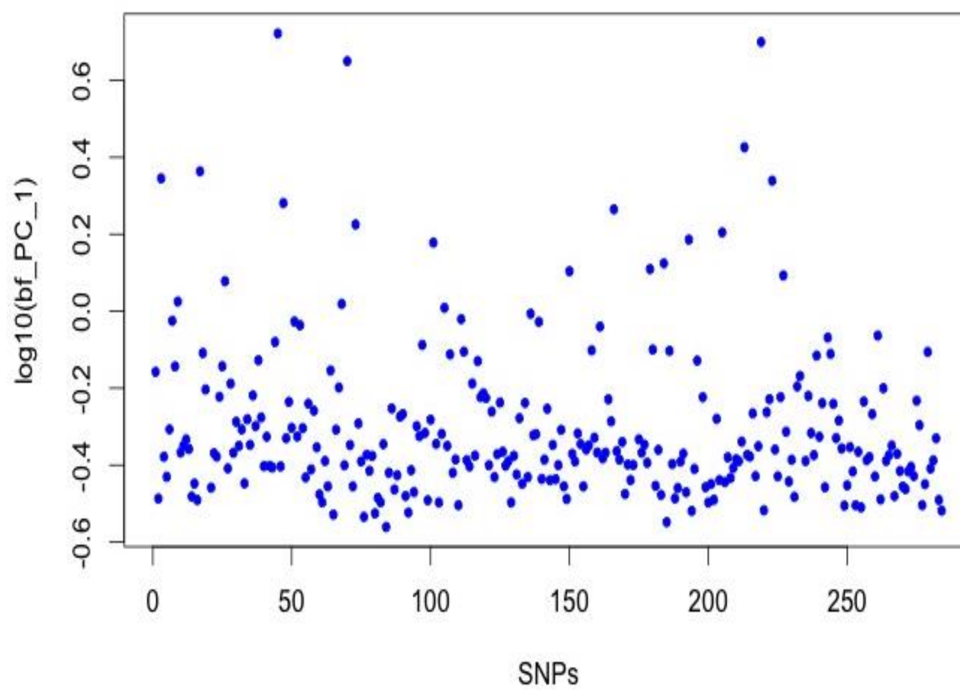
As environmental matrix was used a dataset containing tree PCs values from PCA analyses performed on seasonal environmental dataset, geo-reference data (e.g. longitude, latitude and elevation) and 20 environmental variables from the seasonal dataset. For every variable, Bayes factor (BF) was estimated and totally there were available 26 Bayes factor, which were converted to logarithm of base 10.



C




D



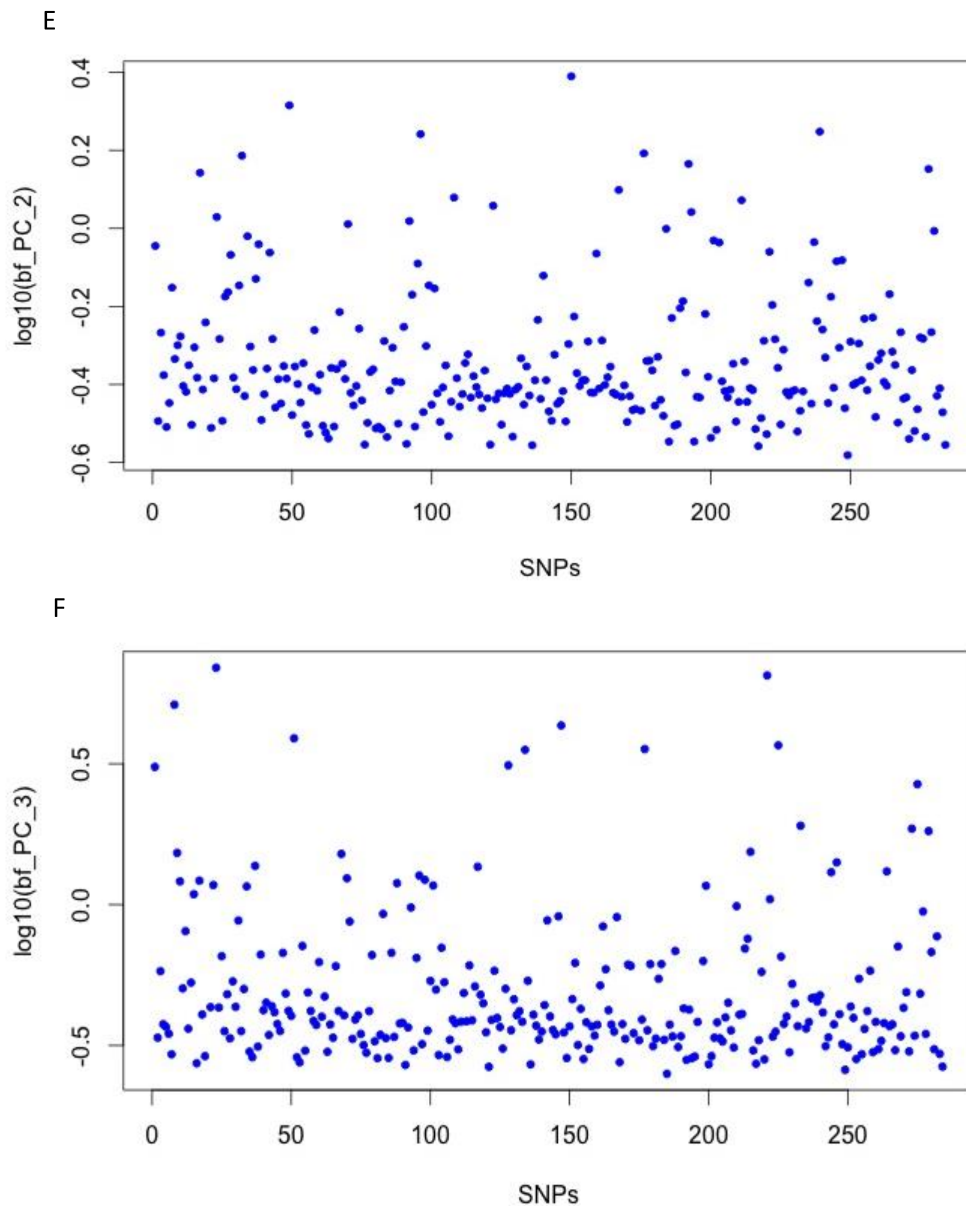



Figure 22: R plots visualization of 280 SNPs and logarithm of Bayes factors for latitude (A); (B) longitude; (C) elevation; (D) PC 1 (the first principal component value); (E) PC 2 (the second principal component value); (F) PC 3 (the third principal component value).

Overall, from total 284 loci, two of them showed strong correlation with geo-climatic variables, indicating a strong evidence of selection. The estimation of the strength for evidence of selection is interpreted based on Jeffrey's scale of evidence described by Jeffrey (1961). Two SNPs indicated strong association to longitude, in particular SNP id 23 (locus name PGWD1-0634) whose logarithm of Bayes factor was equal to 1.29 (Bayes factor equal to 19.6) (Figure 22- B). Another SNP id 134 (locus name PaPHYN_RIII272) showed also strong correlation to longitude, whose logarithm of Bayes factor was equal to 1.07 (Bayes factor equal to 11.72) (Figure 22- B). Regarding other geo-variables (e.g. latitude and elevation), loci indicated weak to moderate correlation (Figure 22- A, C). In addition, all three principal components loaded mainly on the seasonal environmental dataset, did not show to be in strong correlation with loci, when logarithms of Bayes factors in all three cases did not reach 1 which presents a signal for substantial evidence of selection (Figure 22- D, E, F). Beside principal component

values, for all seasonal environmental variables, was estimated a Bayes factor as well. These variables have not been previously used to reduce their own dimensionality via principal component analyses. Two loci indicated strong evidence of selection when for SNP id 23 (locus PGWD1-0634), logarithm of BF was equal to 1.01 (Bayes factor equal to 10.27) showing strong correlation to winter aridity (Figure 23). The second SNP id 96 (PGLM2-0703) whose logarithm of BF was equal to 1.14 (Bayes factor equal to 14.02) was strongly correlated to autumn precipitation (Figure 23).

The locus PGWD1-0634 shared both correlations to longitude and winter aridity. It has originated from Arborea dataset (Canada) and encodes a GMD1 (GDP-D-mannose 4,6-dehydratase 1) and it is involved in the fucose biosynthesis pathway and catalytic and coenzyme binding. The second locus whose correlation was to longitude, was PaPHYN_RIII272 sourced from Uppsala (Sweden) dataset and located in the *Picea abies* partial gene for phytochrome N. The last locus strongly correlated to autumn precipitation was PGLM2-0703 originated from Arborea (Canada) dataset and encoding for mitochondrial substrate carrier family protein.

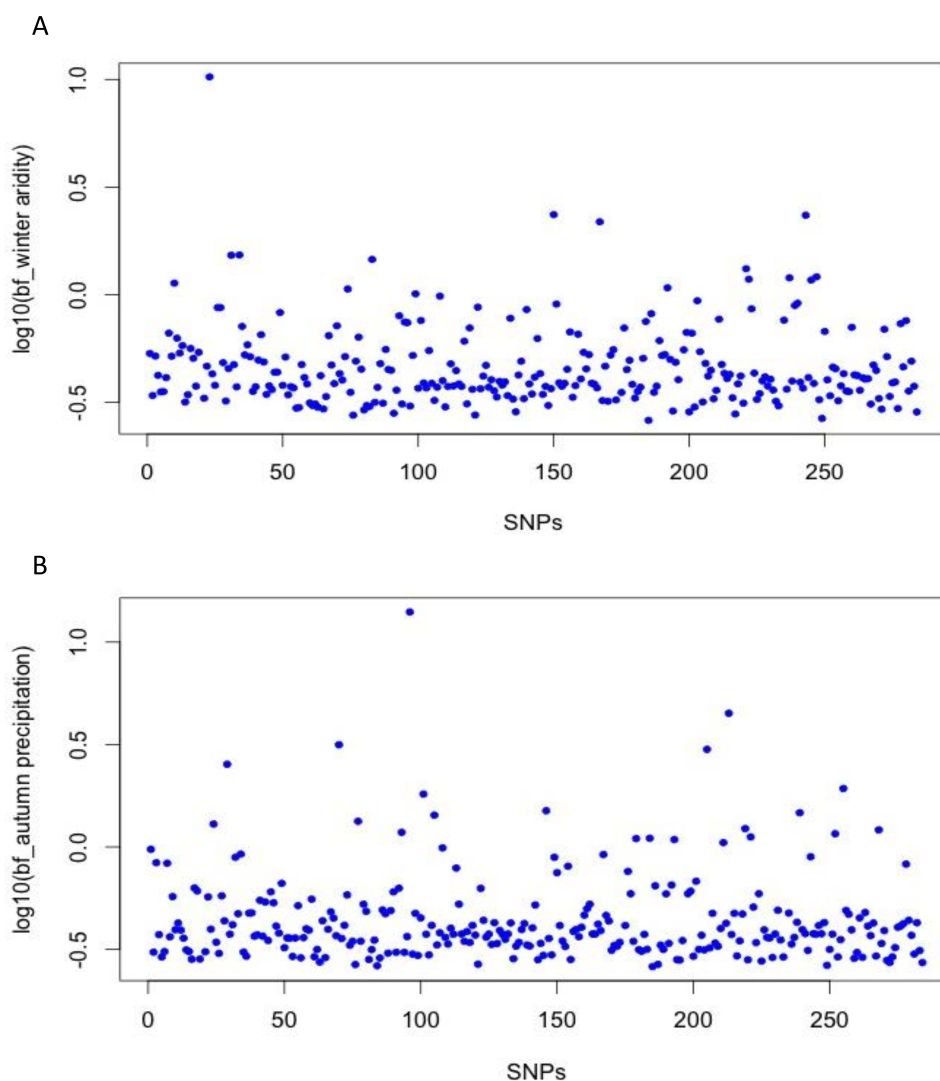



Figure 23: R plots visualizations for correlation between locus id 23 (PGWD1-0634) and logarithm of base 10 of Bayes factors for winter aridity (A) and locus id 96 (PGLM2-0703) and logarithm of base 10 of Bayes factors for autumn precipitation (B).

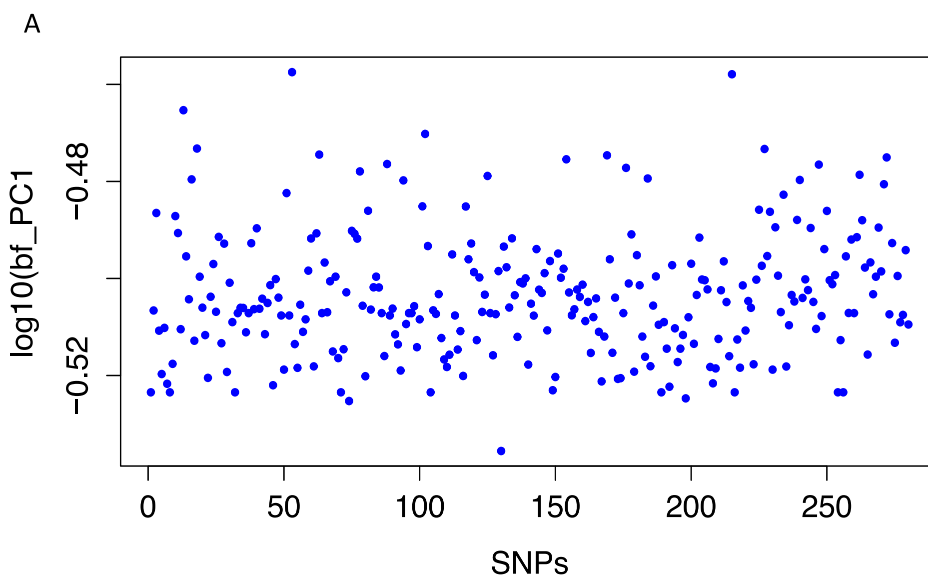
In general, there was no detection of signals for the very strong correlations, indicating that most of SNPs were not very strongly associated with geo-climatic variables. Logarithms of Bayes factors indicated the evidence more in favor of the null model (neutral) over alternative (selective) model. Among all loci included in analyses, three loci appeared to be strongly associated, whose Bayes factor was >10. However, Bayes factor did not attend to be higher than 100 signaling very strong signal of adaptation (Table 16).

Table 16: Summary of SNPs showed to have strong association to geo-climatic variables. Listed are SNPs for which BFs greater than 10 with their functional annotations.

| Variable | BF | SNP | Annotation |
|---|---|---|---|
| Longitude | 19.6 | PGWD1-0634 | GDP-D-mannose 4,6 dehydratase |
| Winter AI | 10.27 | PGWD1-0634 | 1(BT106868) |
| Autumn P | 14.02 | PGLM2-0703 | Mitochondrial substrate carrier protein (BT115596) |
| Longitude | 11.72 | PaPHYN_RIII272 | *P. abies* partial phynrl gene for phytohrome N (JQ970263) |

AI, aridity; P, precipitation; BF, Bayes factor

Individual based approach reveled very weak signal detection of the selection when tested only four PCs (principal component values) from the monthly dataset. For this approach, was decided to include environmental dataset for monthly variables since it has showed slightly the higher proportion of variance (92.4%) than for the seasonal dataset (90.4%). Results revealed that logarithm of base 10 of Bayes factors for all four principal components were very low and ranged from -0.52 up to -0.40 indicating no presence of selection when this approach applied (Figure 24.).



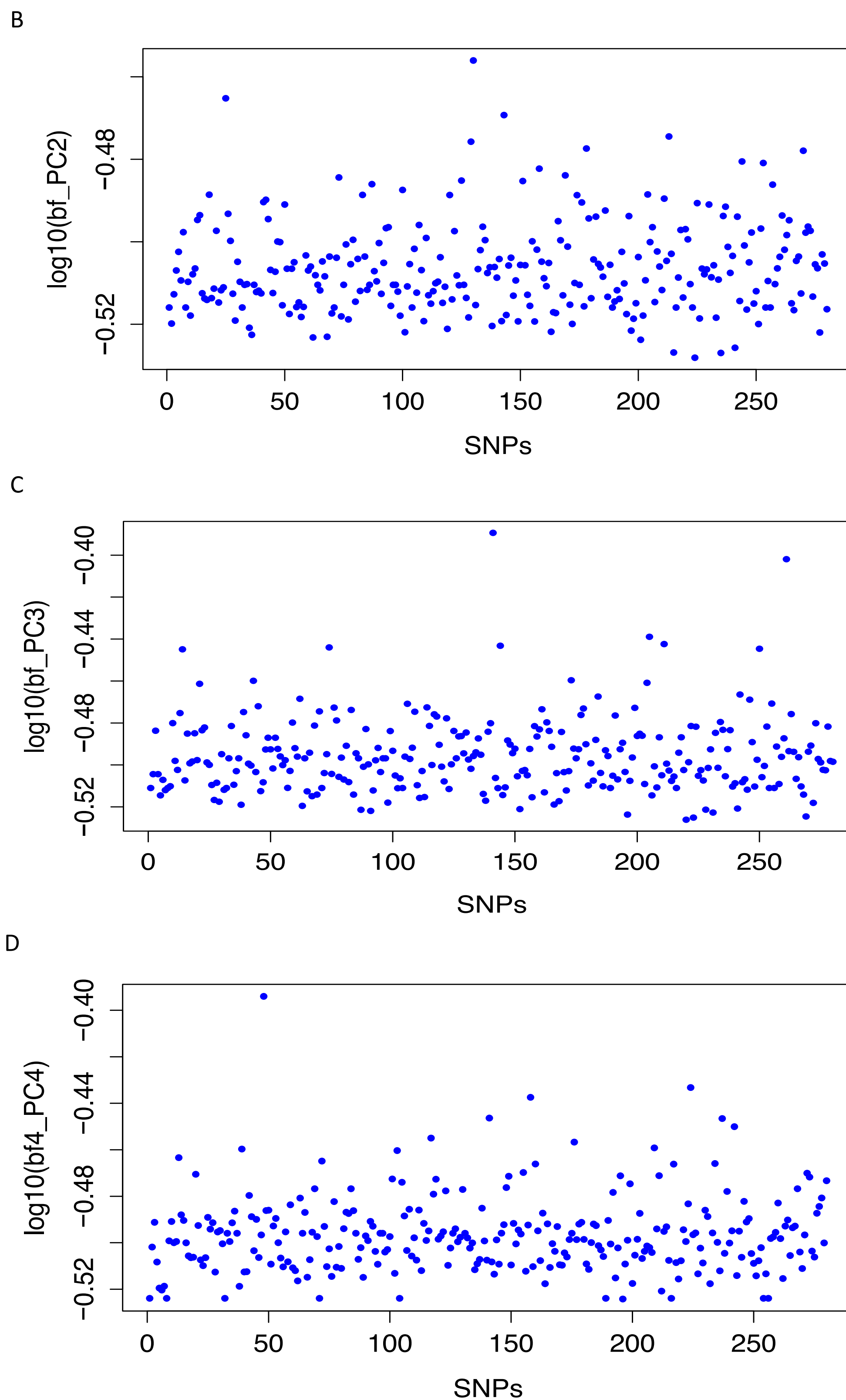



Figure 24: R plots visualization of 280 SNPs and logarithm of Bayes factors for PC 1 (the first principal component value) (A); (B) PC 2 (the second principal component value); (C) PC 3 (the third principal component); (D) PC 4 (the fourth principal component value.

## 4.7 *Spatial analyses detection (Samβada)*

Spatial analyses detection method (Samβada v. 0.4) produced two output files, where for logistic regression analyses, was obtained one unsorted file for constant models (Supplement 8). The first column is the name of the molecular marker combined with allele name. The following columns are Loglikelihood values, the frequency of the marker, estimation of the parameter $\beta_0$ for the logistic model and the error code (default "0" if success). The second obtained file was file for univariate models, which contained list of molecular markers sorted according to G score and WaldScore as well as Loglikelihood values. Next columns were containing measures for the regression (pseudo-$R^2$), followed by AIC (Akaike information criterion) and BIC (Bayesian information criterion). The last columns were placed parameters for regression ($\beta$), one constant parameter and one corresponding to the environmental variable (Figure 2). The second file is the most informative one since it contained the most significant models based on the threshold criterion, which in this particular case was equal to 0.00056=0.01/(280*3*21) where 0.01 is a threshold for significance or Bonferroni's correction threshold; 280 is a number of loci included in analyses and screened for signal of selection, 3 represents three distinct genotypes for each marker and 21 presents a number of environmental variables considered which came from seasonal dataset. However, none of significant model was detected within this threshold.

In addition, another trial has been done when genetic and environmental data have been previously pruned and when a new threshold of significance was created. Pruning of environmental variables was done to keep only those which are not highly correlated since it could lead to an over conservative significance threshold with the Bonferroni correction. After pruning, only six variables remained which did not have high correlation. Pruning of genetic data was based on minor allele frequency when loci were kept if their minor allele frequency is equal or higher than 0.05, retaining only polymorphic loci. Samβada analyses were conducted based on revised dataset by using option BEST saving only the best models corresponding to significance level specified with new threshold applied which was equal to $2.67\ 10^{-5}$ = (0.01/(208*3*6) where 208 presents polymorphic SNPs, 3 presents three distinct genotypes for each loci and 6 represents six environmental variables which have not been highly correlated ("ppt_win"- winter precipitation; "tmax_win" – maximum temperature winter; "ppt_sum" – summer precipitation; "ppt_aut" – autumn precipitation; altitude and "AI spring" – spring aridity). No significant model has been detected either.

## 5. DISCUSSION

### 5.1 *The thesis summary*

This study represents one of the several exploratory studies conducted on adaptive genetic potential in Conifer species. Although other examples of similar studies are available, from these discoveries not very high proportion of novel putative genes has been discovered. The genetic basis of adaptation in forest species, especially in Conifers is still not entirely investigated.

We used a 384 Illumina GoldenGate genotyping approach with single nucleotide polymorphisms (SNPs) to estimate adaptive genetic potential in Norway spruce's population. Totally 400 individuals were genotyped, resulting in successful genotyping of 394 individuals where 6 failed due to the bad quality of genomic DNA used for sequencing. Basic diversity statistics of genotyping dataset indicated that from total 384 loci, a hundred loci failed during genotyping where the most of them are originated from Arborea project designed on white spruce (*Picea glauc*a). The reason of such high number of SNPs, which failed during genotyping, is due to the very little conservation of SNPs even between closely related species, in this case within *Picea abies* and *Picea glauca*. Finally, a total number of polymorphic SNPs were 208, which were included into outlier detection analyses.

Attempt to cluster of individuals into populations did not result with a strong population structure, indicating in particular existence of panmictic population where all individuals are randomly mating or all mixed. In such case, population is a single entity where processes such as mating and movement of individuals are uniform showing no heterogeneity. Within absence of the strong population structure, potentially strong gene flow or the rate of mixing is going on. Norway spruce has an ability of strong pollen and seed dispersal enabling their movement for hundreds of kilometers away or even further via waterways and human or animal distribution.

With no strong population structure ongoing within natural population, outlier detection analyses were individually based oriented. The search for signal of selection resulted in a relatively weak signal detection where the highest Bayes factor was around 20 showing a strong association to longitude whereas association with climatic gradients was common for winter aridity and autumn precipitation among all environmental gradients from the seasonal dataset.

Environmental association analyses to search for correlations between climate variables and single nucleotide polymorphisms (SNPs) across the range sample of loblolly pine, were performed by Eckert *et al*. (2012b). Applications of a Bayesian linear mixed model provided identification of 22 SNPs with very strong signal of selection (Bayes factor >100) being located in genes with functional annotations involved in different mechanisms of plant response (e.g. oxidative stress, cell membrane related, sugar metabolism) to abiotic stress to drought, salt or cold tolerance. Statistical power for detection of loci under the selection was higher when included 1730 loci from 682 individuals of loblolly pine (*Pinus taeda* L.), which were sampled from 54 populations. This genome-wide dataset for loblolly pine was based on the large geographic scale coverage resulted in a discovery of associated loci thanks to population structure, which was accounted within natural populations of loblolly pine.

Another evidence of the local adaptation within conifer species across the Alps, was reported by Mosca *et al*. (2012). In this study single nucleotide polymorphisms (SNPs) were genotyped in 24 to 37 populations from four subalpine conifer species: *Abies alba* Mill, *Larix decidua* Mill, *Pinus cembra* L. and *Pinus mugo* Turra. A total of 3898 individuals were genotyped and tested for a correlation with geo-environmental data. In all four conifer species, was detected population structure ranking from K=5 (*Abies alba*); K=3 (*Larix decidua*), K=4 (*Pinus cembra*) and K=4 (*Pinus mugo*). This study resulted in the successful detection of loci ranging from six in *Larix decidua* to 18 in *Pinus mugo* mostly associated to winter precipitation and seasonal minimum temperature, whereas in *Abies alba*, two loci resulted in association to the seasonal minimum temperature.

A more powerful attempt to detect patterns of local adaptation in forest species when used landscape genomics approach, lies in the high number of loci and individuals. In this study were applied two different detection methods for outlier scans and none of them resulted with strong signal detection. A Bayesian liner mixed model-BayEnv2 (Gunther and Coop, 2013) implements a Bayesian method to compute correlations between allele frequencies and ecological variables into account, whereas Samβada's approach (Stucki and Joost, 2014) integrates logistic regression method to model the probability of the presence of an allelic variant for the polymorphic marker given the environmental conditions. Stucki *et al*. (2014) compared several population-based methods with Samβada analyses and concluded that top detection methods in number of outlier loci are Bayesian linear mixed model (Bayenv2.0) and Samβada often resulting also in a common detection of loci. Although different, these two approaches often provide the highest percentage of detected SNPs. Stucki *et al*. (2014) emphasized the possibility of false positives when used Spatial analyses methods, whereas for Bayesian linear mixed model false positives are less common showing some higher power and greater robustness to population structure (Gunther and Coop, 2013). Although computationally demanded, an advantage of this Bayesian method is the possibility where each individual is treated as a population, allowing to the natural population differentiation test statistics to be calculated without regard of any prior population label (Gunther and Coop, 2013).

Lotterhos and Whitlock (2014) compared several methods to test for selection: FDIST2, BayeScan, FLK and Bayenv2 for the false-positive rates and the power of those four methods to detect loci which differentiated by spatially heterogeneous selection. Older and more widely used methods, such as FDIST and BayeScan showed to have more false-positive rates for loci under either divergent or balanced selection, whereas Bayenv2 and FLK methods showed to be more promised (Lotterhos and Whitlock, 2014).
When identifying outlier loci, it is generally necessary to perform multiple tests to minimize the false positive rate (Perez-Figueroa *et al*., 2010). It is therefore important to validate the detected outlier loci in multiple ways to determine whether or not they are adaptive such as quantitative trait locus mapping, testing for genotyping errors or consideration of their genomic location (Luikart *et al*. 2003).

## 5.2 *Limitations of the study*

As genomic data became more available, genome scans for signatures of adaptation became more common (Tiffin and Ross Ibarra, 2014). Although several studies have been conducted to explore an adaptive genetic potential in conifer species, a small proportion of novel putative genes was discovered. Similar, in case of Norway spruce, a genomic approach with 384 Illumina GoldenGate method provided only a small insight of adaptation based on studies conducted so far.

One of the pioneer attempts to investigate adaptive genetic variation in Norway spruce was done by Scalfi *et al*. (2014). A total of 384 SNPs representing 290 genes covering different cellular metabolic functions, were genotyped at micro and macro-geographic scale (Table 17). To identify outlier loci methods such as testing Fst outliers (BayeScan software v.2.1) and linear regression methods were combined when seven Fst-outliers were detected at macro-geographic scale whereas two outliers were detected at micro-geographic scale by linear regression method. Totally eight genes potentially involved for adaptation, were identified within these two methods when seven genes were identified with BayeScan v.2.1 method (Foll and Gagiotti, 2006) and one gene with linear regression method (Table 18). Di Pierro *et al*. (in preparation) continued to explore an adaptive potential in Norway spruce when totally 826 trees were genotyped with 384 SNPs belonging to 285 putative candidate genes (Table 17). Outlier analyses were performed with BayeScan program v. 2.1 (Foll and Gagiotti, 2008) resulted in two genes discovery. To access the correlation between SNP allele frequencies and climatic variables, a Bayesian generalized linear mixed model was applied (Coop *et al*. 2010) which provided detection of four genes (Table 18). The most recent attempt to investigate adaptive variation was performed when used 392 mother trees of Norway spruce sampled across Alps for 384 SNPs representing 242 genes (Table 17). In this particular approach, Fst outlier analysis with BayeScan v. 2.1 (Foll and Gagiotti, 2008) was performed when none of loci were detected as outliers at 5% significant level. On the other hand, Bayenv analyses resulted in weak signal detection for two loci showing strong correlation to geo-climatic variables when two genes were potentially correlated (Table 18).

Table 17. Norway spruce sampled individuals included for 384 Illumina genotyping assay

| **Number of individuals** | **Sampling design** | **Number of SNPs genotyped** | **Reference** |
|---|---|---|---|
| 300 trees (microscale) | 6 populations (25 ind/site) | 384 SNPs | Scalfi *et al*. 2014 |
| 546 trees (macroscale) | 27 populations (15-24 ind/site) | 384 SNPs | Scalfi *et al.* 2014 |
| 826 trees | 24 populations (25-65 ind/site) | 384 SNPs | Di Pierro *et al.* (in preparation) |
| 392 trees | 2-4 ind/site (mean) | 384 SNPs | Calic *et al.* (unpublished) |

Table 18. Summary of candidate genes potentially involved in adaptation process when different outlier detection methods performed.

| Reference | Methods | Number of genes | Total |
|---|---|---|---|
| Scalfi *et al*. 2014 | BayeScan v.2.1 | 7 | 8 |
| | Linear regression method | 1 | |
| Di Pierro *et al*. (in prep.) | BayeScan v.2.1 | 2 | 6 |
| | Bayenv | 4 | |
| Calic *et al*. (unpublished) | Bayescan v.2.1 | 0 | 3 |
| | Bayenv | 3 | |
| | Samβada | 0 | |

Overall, among these three studies, the most common method of detection was BayeScan v.2.0, that is using allele frequencies between populations to identify candidate loci. In particular, it resulted in detection of nine candidate genes within studies both conducted by Scalfi *et al*. (2014) and Di Pierro *et al*. (in preparation) but not providing any detection of outlier loci in the case of the last study (Calic *et al*. unpublished). Both Scalfi *et al*. (2014) and Di Pierro *et al.* (in preparation) studies provided estimation of population structure of the sampled Norway spruce population. For instance, Scalfi *et al*. (2014) scanned the genome of Norway spruce on macro scale when 6 populations (25 individuals/site) were included and 27 populations (15-24 individuals/site) were included in macro-geographic scale. Population structure was revealed only on macro scale indicating four clusters with Bayesian cluster analyses –STRUCTURE (Pritchard *et al*. 2000). Di Pierro *et al*. (in preparation) also estimated population structure when detected four genetic clusters. On the other hand, the study by Calic *et al*. unpublished, estimated an existence of panmictic population indicating a strong gene flow or a rate of mixing going on. Norway spruce is known to have a strong pollen and seed dispersal enabling their movement for hundreds of kilometers away or even further via waterways and human or animal distribution. An efficient detection of outlier loci is usually fruitful when the population structure is stronger. Other genomic approaches are available such as Sambada method, which integrates logistic regression method to model the probability of presence of an allelic variant for the polymorphic marker given the environmental conditions (Stucki *et al*. 2014). Although not requiring any population structure of sampled natural population, its disadvantage is the possibility of existence of false positives.

A 384 Illumina GoldenGate method provided only a small percentage of genes identified to be responsible for adaptation. In particular, less than 3% of genes have been discovered from the total number of genes included (Table 19). The highest number of genes (8) was discovered by Scalfi *et al*. (2014) when 846 trees were included on both micro and macro geographic scale. Di Pierro *et al*. (in preparation) resulted in discovery of six genes when 826 individuals were genotyped, where Calic *et al*. (unpublished) resulted in three genes relatively correlated to geo-environmental variables when included 392 individuals.

As outcome from each study on N. spruce, a pairwise combination of genes 8-6-3 has been discovered (Table 19). However, among genes potentially adapted, there was no evidence of repeatability (Table

20). Discovered genes are putatively involved in various biological processes (e.g. plant defense, photosystem regulation, sugar or lipid metabolism etc.) when each study resulted with uncommon candidate gene annotations.

Table 19. The percentage (%) of discovered genes from the total number of genes screened for signal of selection

| **Number of genes** | **Genes responsible for adaptation** | **Percentage (%)** | **Reference** |
|---|---|---|---|
| 290 | 8 | 2.76% | Scalfi *et al*. (2014) |
| 285 | 6 | 2.11% | Di Pierro *et al*. (in prep.) |
| 242 | 3 | 1.24% | Calic *et al. (unpublished)* |

Limitations in discoveries of novel candidate genes are related to both numbers of loci genotyped as well as sampling design of the population. An approach based on 384 SNPs did not approve a high detection power for adaptive loci in studies conducted so far in N. spruce. The discovery of genes potentially adapted, only tells about a small proportion of genome which has been screened and that there is still much more to be discovered. To provide this, GWAS (genome wide association studies) certainly could provide more detection power in search for adaptation in N. spruce, whose genome is estimated to 20 gigabases (GB) (Nysted *et al*. 2013).

On the other hand, if sampling provides an estimation of the population structure, the probability to detect more outlier loci is higher. Besides the number of loci, the genetic diversity plays important role when selection ongoing. With the presence of homogeneous population, the signal of selection will be lost or with weak signal detection. For instance, Di Pierro *et al*. (in preparation) and Scalfi *et al*. (2014) have observed a sufficient genetic diversity within natural populations, which enabled the discovery of outlier loci. On the other hand, Calic *et al*. (unpublished) has observed a homogenous population with very low Fst values, indicating that all individuals are uniform and in such case there was no strong signal of selection to be detected.

Table 20. Summary of gene candidates potentially involved in adaptation generated with different outlier methods confirms no repeatability existence

| **Gene classification** | **Methods** | **Reference** |
|---|---|---|
| Plant defense responses | Fst_outlier detection | Scalfi *et al.*2014 |
| Lipid metabolism | | |
| Sucrose synthase protein in *Pinus halepensis* | | |
| F box protein | | |
| Sugar metabolism | | |
| Various biological functions | | |
| Nucleic acid-binding protein | | |
| Pentatricopeptide (PPR) repeat-containing protein of *A. thaliana* | Regression method | |
| Poly-adenilate binding domain | Fst outlier detection | Di Pierro *et al*. (in preparation) |
| Putative galactokinase | | |
| Gene encoding for SNF2 proteins | Bayenv | |
| Oxygen-evolving protein 1 (photosynthesis) | | |
| Phosphoenolpyruvate carboxynase 2 (Gluconeogenesis) | | |
| NADP(P)-linked oxidoreductase | | |
| / | Fst outlier detection | Calic *et al*. (unpublished) |
| GMD1(GDP-D-mannose 4,6-dehydratase 1) | Bayenv | |
| Phytochrome N | | |
| Mitochondrial substrate carrier family protein | | |
| / | Spatial analyses method | |

## 5.3 *Future perspectives*

Forest tree species have not been extensively domesticated as crop species and natural populations are still abundant (Tsumura *et al.* 2012). The new insights into the adaptive mechanisms are possible by identifying and studying adaptive genes in these natural populations (Neale and Savolainen, 2004, Neale and Ingvarsson, 2008). Since forest tree species are facilitated by relatively low genetic differentiation between populations on average, are widely distributed in different environments and have relatively large population sizes (Tsumura *et al.,* 2012). As a consequence, the selective pressures acting on a population in a given environment tend to result in adaptation by selecting for a relatively small number of genotype changes at a few specific loci (Pelgas *et al*., 2011).

Current population genomic methods are not well posed to identify adaptive loci (Berg and Coop, 2014). However, recent advances in population genetics allowed for genome-wide identification of individual recent selective events by identifying unusually large allele frequency differences among populations and environments o by detecting effects of these events on linked diversity (Nielsen *et al*. 2005). Berg and Coop (2014) explained limitations of these population genomic method approaches correlated to the possibility only to identify traits under the selection where an individual allele has a large and/or sustained effect on fitness. As explained, when selection acts on a phenotype that is underwritten by a large number of loci, the response at any given locus is expected to be modest and the signal manifests in shift in allele frequency across many loci (Kremer and Corre, 2011). This signal would be very weak at the level of individual loci that it may be impossible to identify it against the genome-wide background without a specific annotation of which sites are the targets of selection on a give trait (Pritchard *et al*. 2010).

The genome wide association studies are more powered when available large set of individuals enabling mapping of many small effects alleles associated with phenotypic variation down to the scale of linkage disequilibrium in the population (Berg and Coop, 2014). The development and application of GWAS has provided an identification of thousands of loci associated with a wide range of traits and in the same time confirming a polygenic view of phenotypic variation (Visscher *et al*. 2012). GWAS is providing a powerful way to identify the signal of adaptation in polygenic traits.

Additionally, more intention should be focused on genotype to phenotype approach, which has proved to be informative on adaptive genetic potential in forest species. Forest trees are difficult to phenotype since they are long lived and sessile organisms, however an improvement in phenotyping is expected to provide more informative, precise and standardized high-throughput phenotyping technologies (Neale and Kremer, 2011). In addition, a good combination of landscape genomics and common garden experiments could provide more insight into the adaptive potential of forest species.

## References

Aitken, S. N. (2004). Genecology and Adaptation of forest trees. Elsevier *Encyclopedia of Forest Sciences Article* 86, pp. 97-2004.

Aitken, S.N., Yeaman, S., Holliday, J.A., Wang, T. and Curtis-McLane S. (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, vol.1, pp.95-111.

Alberto F. J., Aitken S. N., Alia R., Gonzalez-Martinez G.C., Hanninnen H., Kremer A., Lefevre F., Lenormand T., Yeaman S., Whetten R., Savolainen O. (2013). Potential for evolutionary responses to climate change-evidence from tree populations. *Global Change Biology*, vol.19: 1645-1661.

Allen C.D. (2009). Climate-induced forest dieback: an escalating global phenomenon? *Unasylva*, vol. 60: 231-232.

Allen, C.D., Macaldy, A.K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D.D., Hogg, E.H., Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, Jong-Hwan, Allard, G., Running, W., Semerci, A., Cobb, N. (2010). A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management*, vol.259. pp 660-684.

Anderson G. B. Principal component analyses in R. An Examination of the different functions and methods to perform PCA. Tutorial available online: http://filebox.vt.edu/users/gba/ePortfolio/Public/AndersonGB_PCAinR.pdf

Austerlitz F., Dick C.W., Dutech C., Klein E., Oddou-Muratorio S., Smouse P.E. and Sork V.L. (2004). Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, 13(4): 937-954.

Berg E.E., Henry J.D., Fastie C.L., De Volder A.D., Matsuoka S.M. (2006). Spruce beetle outbreaks on the Kenai Peninsula Alaska and Kluane National Park and Reserve, Yukon Territory: relationship to summer temperatures and regional differences in disturbance regimes. *Forest Ecology and Management*, vol. 227: 219-232.

Berry A. and Kreitman M. (1993). Molecular analysis of an allozyme cline: Alcohol-Dehydrogenase in Drosophila melanogaster on the East coast of North America. *Genetics*. vol. 134: 869-893.

Bigler C, Braker O.U., Bugmann H., Dobbertin M., Rigling A. (2006). Drought as an inciting mortality factor in Scots pine stands of the Valais, Switzerland. *Eco-systems*, vol.9: 330-343.

Chen J., Kallman T., Ma X., Gyllenstrand N., Zaina G., Morgante M., Bousquet J., Eckert A., Wegrzyn J., Neale D., Lagercrantz U. and Lascoux M. (2012). Disentangling the Roles of History and Local Selection in Shaping Clinal Variation of Allele Frequencies and Gene Expression in Norway Spruce (*Picea abies*). *Genetics*, vol.191, pp.865-881.

Christensen, J.H., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., Jones, R., kolli, R.K., Kwon, W.-T., Laprise, R., Magnana Rueda, V., Mearns, L., Menendez, C.G., Raisanen, J., Rinke, A., Sarr, A., Whetton, P. (2007). Regional climate projections. In: Solomon, S., et al. (Eds), Climate change, Cambridge University Press, Cambridge, United Kingdom/New York, NY.

Christersson L., Fircks H. (1990). Frost and winter desiccation as stress factors. Aquilo, *Ser. Bot*. 29:13-19.

Coop G., Witonsky D. Di Rienzo A. and Pritchard J.K. (2010). Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, vol. 185: 1411-1423.

Davis, M.B., Shaw, R.G. and Etterson, J.R. (2005). Evolutionary responses to changing climate. *Ecology*, 86(7). pp1704-1714.

Daehlen AG. Johnsen O. and Kohmann K. (1995). Hostfrostherdighent hos unge granplanter fra norske provenienser og iroplantasjer. *Rapport fra Skogforsk* 1/95, 24 pp.
De La Torre A. R., Roberts D. R. and Aitken S. A. (2014). Genome-wide admixture and ecological niche modeling reveal the maintenance of species boundaries despite long history of interspecific gene flow. *Molecular ecology*, vol.23: 2046-2059.

Debreczy, Zs. and I. Rácz, (2011). Conifers around the world. vol.1- DendroPress Ltd., Budapest. pp 1089.

Di Pierro E. A., Mosca E., Rocchini D., Binelli, Neale D. B. and La Porta N. (unpublished). Population structure and adaptive genetic variation to climate in natural populations of Norway Spruce (*Picea abies* [L.] Karst).

Dobbertin M, Rigling A. (2006). Pine mistletoe (*Viscum album* spp. *austriacum)* contributes to Scots pine (*Pinus sylvestris*) mortality in the Rhone valley of Switzerland. *Forest Pathology*, vol.36: 309-322.

Earl, Dent A. and vonHoldt, Bridgett M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, vol. 4: 359-361.

Eckert A. J., Bower A. D., Gonzales-Martinez S. C., Wegzyn J. L., Coop G., Neale D.B. (2010b). Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology*, vol. 19, pp.3789-3805.

Eckert A.J., van Heerwaarden J., Wegrzyn J.L., Nelson D.C., Ross-Ibarra J., Gonzalez-Martinez S.C. and Neale D.B. (2010a). Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (*Pinus taeda* L., Pinaceae). *Genetics*, 185: 969-982.

Ellenberg H. (1978). Vegetation Mitteleuropas mt den Alpen in okologischer Sicht. Verlag Eugen Ulmer, Stuttgart.

Evanno G., Regnaut S. and Goudet J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, vol. 14: 2611-2620.

Everitt B. and Hothorn T. (2011). An Introduction to Applied Multivariate Analysis with R. Springer Science+Business Media. pp: 61-102.

Fan J.B., Oliphant A., Shen R., Kermani B.G., Garcia F., Gunderson K.L., Hansen M., Steemers F., Butler S.L., Deloukas P., Glaver L., Hunt S., McBride C., Bibikova M., ubano T., Chen J., Wickham E., Doucet D., Chang W., Campbell D., Zhang B., Kruglyak S., Bentley D., Haas J., Rigault P., Zhou L., Stuelpnagel J., Chee M.S. (2003). Highly parallel SNP Genotyping. Volume LXVIII, Cold Spring Harbor Laboratory Press.

Farjon, A. and D. Filer. (2013). An Atlas of the World's Conifers. Brill.

Foll M. (2010). BayeScan v2.0 User Manual available online: http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.0_manual.pdf

Foll M. and Gaggiotti O. (2008). A genome-Scan Method to Identify Selected loci Appropriate for both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* (180): 977-993.

Forester B., Meier F., Brandli U.B. (2008). Deutlicher Ruckgang der Fichten im Mittelland. Vorratsabbau-auch durch Sturm und kafer. *Wald Holz,* vol. 89(3): 52-54.

Forsythe W. C., Rykiel Jr. E.J, Stahl R. S., Wu Hsin-I, Schoolfield R. M. (1995). A model comparison for daylenght as a function of latitude and day of year. *Ecological Modelling*, vol. 80: 87-95.

Futuyma D.J. (2005). Evolution. Sinauer Associates, Inc. Sunderland, Massachusetts U.S.A. pp.247-267.

Goudet J. (2014). Package "Hierfstat" in R. Tutorial available online: http://www.r-project.org, http://www.unilch/popgen/softares/hierfstat.htm

Grant, V. (1963). The Origin of Adaptations. Columbia University press. New York and London. pp 118.

Green P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*. vol. 82: 711-732.

Grivet D., Sebastiani F., Alia R., Bataillon T., Torre S., Zabal-Aguirre M., Vendramin G.G. and Gonzalez-Martinez C.S. (2011). Molecular Footprints of Local Adaptation in Two Mediterranean Conifers. *Mol. Biol. Evol*. vol. 28: 101-106.

Gunther T. and Coop G. (2013). Robust Identification of Local Adaptation from Allele frequencies. *Genetics*, vol. 195: 205-220.

Gyllenstrand N., Clapham D., Kallman T. and Lagercrantz U. (2007). A Norway Spruce Flowering Locus T Homolog is implicated in Control of Growth Rhythm in Conifers. *Plant Physiology*, vol. 144: 248-257.

Harrison L.C. (1960). Sun, Earth, Time and Man. Rand McNally & Co., Chicago, IL.

Hijmans R. J, Cameron S. E., Parra J.L., Jones P.G., Jarvis A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatology*, vol.25: 1965-1978.

Holderegger R., Buehler D., Gugerli F. and Manel S. (2010). Landscape genetics of plants. *Trends in Plant Science*, vol. 15, n.2, pp. 675-683.

Holderegger R., Herrmann D., Poncet B., Gugerli F., Thuiller W., Taberlet P., Gielly L., Rioux D., Brodbeck S., Aubert S. and Manel S. (2008). Land ahead: using genome scans to identify molecular marker of adaptive relevance. *Plant Ecology and Diversity*, vol.1, n.2, pp. 273-283.

Holland S.M. (2008). Principal components analyses (PCA). Online tutorial available: http://strata.uga.edu/software/pdf/pcaTutorial.pdf

IPCC (2007b). Climate change: impacts, adaptation and vulnerability. Parry M.L., Canziani O.F., Palutikof J.P., van der Linden P.J., Hanson C.E. (eds). Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, 976pp.

Jeffreys H. (1935). "Some Tests of Significance, Treated by the Theory of Probability," *Proceedings of the Cambridge Philisophy Society*, vol. 31: 203-222.

Jeffreys H. (1961). "*Theory of Probability*" (3rd ed.), Oxford, U.K.: Oxford University Press.

Jolliffe I. T. (2002). Principal component analyses, Second Edition. Springer. pp: 1-408.

Joost S., Bonin A., Bruford W., Despres L., Conord C., Erhardt G. and Taberlet P. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach adaptation. *Molecular Ecology*, vol. 16: 3955-3969.

Kawecki T. and Ebert D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, vol. 7: 1225-1241.

Kellomaki S., Rouvienen I., Peltola H., Strandman H., Steinbrecher R. (2001). Impact of global warming on the tree species composition of boreal forests in Finland and effects on emission of isoprenoids. *Glob. Change Biol*. vol. 7:531-44.

Kemper KE, Saxton SJ, Bolormaa S, Hayes BJ, Goddard ME (2014). Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics*, vol.15: 1-14.

King J. R. and Jackson D. A. (1999). Variable selection in large environmental data sets using principal components analyses. *Environmetrics*, vol 10: 67-77.

Koski V., Skroppa T., Paule L., Wolf H., Turek J. (1997). Technical guidelines for genetic conservation of Norway spruce (*Picea abies* (L.) Karst). International Plant Genetic Resource Institute, Rome, Italy. European Forest Genetic Resources Programme (EUFORGEN).

Koskela J., Buck A. and du Cross T. (2007). Climate change and forest genetic diversity. Implications for sustainable forest management in Europe. European Forest Genetic Resources Program. EUFORGEN.
Kremer, A. (2007). How well can existing forests withstand climate change? In Climate change and forest genetic diversity: implications for sustainable forest management in Europe. Eds. Koskela, J., Buck, A. and Tessier du Cros, E.. Biodiversity International, Rome. Italy, pp3-17.

Kremer, A. (2010). Evolutionary responses of European oaks to climate change, *Irish Forestry*, pp53-65.

Kremer A. and Le Corre V. (2011). Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity*, vol.108: 375-385.

Kremer, A., Le Corie, V., Petit, R.J. and Ducousso, A. (2010). Historical and contemporary dynamics of adaptive differentiation in European oaks. In: Molecular Approaches in Natural Resource Conservation (eds DeWoody, A., Bickham, J., Michler, C., Nichols, K., Rhodes, G. and Woeste, K.). Cambridge University Press, Cambridge, pp 101-117.

Kremer, A., Ronce, O., Robledo-Arnuncio, J.J., guillaume, F., Bohrer, G., Nathan, R., Bridle, J.R., Gomulkiewicz, R., Klein, E.K., Rittland, L., Kuparinen, A., Gerber, S., Schueler, S. (2012). Long distance gene flow and adaptation of forest trees to rapid climate change. *Ecology letters*.

Kurz W.A., Dymond C.C., Stinson G., Rampley G.J., Neilson E.T. Caroll A. L., Ebata T., Safranyik L. (2008a). Mountain pine beetle and forest carbon feedback to climate change. *Nature*, vol.452: 987-990.

Lotterhos K. and Whitlock M.C. (2014). Evaluation of demographic history and neutral parameterization on the performance of Fst outlier tests. *Molecular Ecology*, vol.23: 2178-2192.

Martinez-Vilalta J, Pinol J. (2002). Drought-induced mortality and hydraulic architecture in pine populations of the NE Iberian Peninsula. *Forest Ecology and Management*, vol.161: 247-256.

Mikola J. (1982). Bud-set phenology as an indicator of climatic adaptation of Scots pine in Finland. *Silva Fenn*. vol.16:178-184.

Modrzynski J. (2007). Outline of ecology. In: Tjoelker M., G., Boratynski A., Bugala W. (2007). Biology and Ecology of Norway Spruce. *Forestry Sciences*, vol.78, pp. 195-253.

Mosca E., Eckert A. J., Di Pierro E. A., Rocchini D., La Porta N., Belletti P. and Neale D. B. (2012). The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular Ecology*, pp.1-16.

Mullin, T.J., Andersson B., Bastien J.C., Beaulieu R. D., Burdon W. S., Dvorak J.N., King J.N., Kondo T., Krakowski J., Lee S.E., McKeand L., Paques A., Raffin J, Russell T., Skroppa T, Stoehr M. and Yanchuk A. (2011). Economic importance, breeding objectives and achievements. Pp. 40-127. *In*: Plomion C., Bousquet J. and Kole C. (eds). Genetics, Genomics and Breeding of Conifers Trees. Edenbridge Science Publishers and CRC Press, New York.

Neale DB, Ingvarsson PL (2008). Population, quantitative and comparative genomics of adaptation in forest trees. *Curr Opin Plant Biol* 11:1-7

Neale D. and Kremer A. (2011). Forest tree genomics: growing resources and applications. *Nature reviews. Genetics*. 111-122.

Neale DB, Savolainen O. (2004). Association genetics of complex traits in conifers. *Trends Plant Sci*, vol. 9:325-330.

Nebe W. (1968). Uber Bezinehungen zwischen Klima und Wachstum der Fichte (*Picea abies* L.) in inherem europeischen Verbreitungsggebiet. Arch. Forstw.17:1219-1238.

Neteler M. (2010). Estimating Daily Land Surface Temperatures in Mountainous Environments by Reconstructed MODIS LST Data. *Remote sensing*, vol. 2: 333-351.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, et al. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research*, vol.15: 283-292.
Nystedt B., Street N.R., Wetterborn A., Zuccolo A., Lin Yao-Cheng, Scofield D.G., Vezzi F., Delhomme N., Giacomello S., Alezeyenko A., Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hallman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Kaller M, Luthman J, Lysholm F, Niittyla T, Olson A, Rilakovic N, Ritland C, Rossello J.A., Sena J, Svensson T, Talavera-Lopez C, Theiβen G, Tuominenn H, Vanneste K, Zci-Qiang W, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R., Hvidsten T.R., De Jong P., Mackay J., Morgante M., Ritland K., Sundberg B., Lee Thompson S, Van de Peer Y., Andresson B., Nilsson O., Ingvarsson P.K., Lundeberg J. and Jansson S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, pp.1-6.

Obminski Z. (1977). Ogolny zarys ekologii. In: Bialobok S. (ed). Swierk pospolity, Picea abies (L.) Karst. Nasze drzewa lesne. PWN, Warszawa-Poznan 5: 332-371.

Obminski Z. (1947). Botanika Lesna z krotkim zaryem botaniki ogolneij. Wyd. Spoldzielni Las, War-Szawa. 112 pp.

Pelgas B., Bousquet J, Meirmans PG, Ritland K, Isabel N. (2011). QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. *BMC Genomics*, vol.12: 145.

Penuelas, J, Lloret F., Montoya R. (2001). Severe drought effects on Mediterranean woody flora in Spain. *Forest Science*, vol.47: 214-218.

Pritchard J.K., Stephens M. and Donnelly P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, vol. 155: 945-959.

Pritchard JK, Pickrell JK and Coop G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps and polygenic adaptation. *Current Biology*, vol.20: R208-15.

Prunier J., Laroche J., Beaulieu J. and Bousquet J. (2011). Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology*, vol. 20: 1702-1716.

Rani A., Singh K., Sood P., Kumar S. and Ahuja P.S. (2009). *p-Coumarate:CoA* ligase as a key gene in the yield of catechins in tea [*Camellia sinensis* (L.) O. Kuntze]. *Funct. Integr Genomics*, vol 9: 271-275.

Savolainen, O., Bokma, F., Knurr, T., Karkkainen, K., Pyhajarvi, T., Wachowiak, W. (2007). Adaptation of forest trees to climate change. In: Koskela, J., Buck, A. and Teisser du Cros, E., editors. Climate change and forest genetic diversity: Implications for sustainable forest management in Europe. Bioversity International, Rome, Italy. Pp. 19-30.

Savolainen, O., Lascoux, M. and Merila, J. (2013). Ecological genomics of local adaptation. *Nature reviews|Genetics*, vol.14, pp.807-820.

Scalfi M., Mosca E., Di Pierro E.A., Troggio M., Vendramin G.G., Sperisen C., La Porta N. and Neale D. B. (unpublished). Interplay between demographic history and selection in shaping genetic diversity in Norway spruce at the micro- and macro-geographic scale.

Schmidt-Vogt H. (1977) (ed). Die Fichte. 1. Taxonomie, Verbreitung, Morphologie, Okologie, Waldesellschaften. Parey, Hamburg u. Berlin. 647 pp.

Schmidt-Vogt H. (1989). Die Fichte. II/2 – Krankheiten. Schaden. Fichtensterben. Verlag Paul Parev, Hamburg and Berlin. 607 pp.
Schoolfield R. (1982). Expressing daylenght as a function of latitude and Julian date. Unpublished.

Seager, R., ting, M., Held, I., Kushnir, Y., Lu, J., Vecchi, G., Huang, H.-P., Harnik, N., Leetmaa, A., Lau, N.-C., Li, C., Velez, J., Naik, N. (2007). Model projections of an imminent transition to a more arid climate in southwestern North America, *Science* 316, 1181-1184.

Skroppa T. and Johnsen O. (2000). Patterns of adaptive genetic variation in forest tree species; the reproductive environment as an evolutionary force in *Picea abies*. In: Matyas, C. (ed): Forest genetics and sustainability, pp. 49-58. Klumer, Dordrecht.

Smith, S.A. and Donoghue, M.J. (2008). Rates of Molecular Evolution are linked to Life History in Flowering Plants. *Science*, vol.322, pp86-89.

Solberg S. (2004). Summer drought: a driver for crown condition and mortality of Norway spruce in Norway. *Forest Pathology*, vol.34: 93-107.

Spiecker, H. (2000). Growth of Norway Spruce (*Picea abies* [L.] Karst.) under changing Environmental Conditions in Europe, in Klimo, E., Hager, H. and Kulhavy, J. (eds) (2000). Spruce Monocultures in Central Europe-Problems and Prospects. EFI Proceedings, 33, pp11-26.

Stewart, W.N. and G.W. Rothwell (1993). Paleobotany and the evolution of the plants. Second edition, Cambridge University Press, Cambridge, UK.

Stucki S. and Joost S. (2014). Samβada: User manual. Available online: http://lasig.epfl.ch/sambada.

Stucki S., Orozco-terWengel P., Bruford M.W., Colli L., Masembe C., Negrini R., Taberlet P., Joost S. and the NEXTGEN Consortium. (2014). High performance computation of landscape genomic models integrating local indices of spatial association.
Cornell University Library. Available online: http://arxiv.org/pdf/1405.7658v1.pdf

Thornthwaite CW. (1948). An approach toward a rational classification of climate. *Geographical Review*, vol. 38: 55-94.

Tiffin P. and Ross-Ibarra J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology and Evolution*. pp.1-8.

Tsumura Y, Uchiyama K, Moriguchi Y, Ueno S, and Ihara-Ujino T. (2012). Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, *Cryptomeria japonica*. *Heredity*. vol.109:349-360.

Visscher PM, Brown MA, McCarthy MI, Yang J. (2012). Five Years of GWAS Discovery. *The American Journal of Human Genetics*, vol.90: 7 -24.

Wagner A., Donaldson L., Kim H., Phillips L., Flint H., Steward D., Torr K., Koch G., Schmitt U. and Ralph J. (2009). Suppression of 4-Coumarate-CoA Ligase in the Coniferous Gymnospersm *Pinus radiata*. *Plant Physiology*. vol. 149: 370-383.

Wright S. (1931). Evolution in Mendelian populations. *Genetics,* vol. 16: 97-159.

Supplement 1: List of mother trees sampled across the European Alps with geo-reference data provided. Total number of genotyped individuals is 392 where 205 individuals from Italy-Slovenia-Austria, 107 individuals from Switzerland and 80 individuals from Austria.

| ID | COUNTRY | LATITUDE | LONGITUDE | ELEVATION |
|---|---|---|---|---|
| **1** | SLOVENIA | 45.85535 | 14.106136 | 779 |
| **2** | SLOVENIA | 45.558689 | 14.442744 | 1142 |
| **3** | SLOVENIA | 46.413892 | 14.330322 | 1278 |
| **4** | ITALY | 45.76277034 | 10.87599277 | 1555 |
| **5** | ITALY | 45.76253831 | 10.87548852 | 1570 |
| **7** | ITALY | 45.76297991 | 10.87557435 | 1555 |
| **8** | ITALY | 46.13353065 | 10.80138445 | 1640 |
| **9** | ITALY | 46.14412432 | 10.80221057 | 1670 |
| **10** | ITALY | 46.14497173 | 10.8034122 | 1670 |
| **11** | ITALY | 46.14508323 | 10.80446362 | 1670 |
| **12** | ITALY | 46.1430985 | 10.80311179 | 1670 |
| **13** | ITALY | 46.358539 | 11.49359822 | 1820 |
| **14** | ITALY | 46.35927203 | 11.49029374 | 1820 |
| **15** | ITALY | 46.36035306 | 11.48886681 | 1820 |
| **16** | ITALY | 46.48599804 | 11.15234613 | 1400 |
| **17** | ITALY | 46.48604975 | 11.15306497 | 1400 |
| **18** | ITALY | 46.4862492 | 10.30208588 | 1370 |
| **19** | ITALY | 46.46715756 | 10.2636981 | 1515 |
| **20** | ITALY | 46.46745315 | 10.26337624 | 1515 |
| **21** | ITALY | 46.46718712 | 10.26352644 | 1515 |
| **22** | ITALY | 46.46972912 | 10.26279688 | 1525 |
| **23** | ITALY | 46.32482361 | 9.502229691 | 900 |
| **24** | ITALY | 46.32566823 | 9.509289265 | 900 |
| **25** | ITALY | 46.32586086 | 9.509589672 | 900 |
| **26** | ITALY | 46.32531261 | 9.508559704 | 800 |
| **27** | ITALY | 46.32476434 | 9.486393929 | 860 |
| **28** | ITALY | 44.18589693 | 7.26998806 | 1590 |
| **29** | ITALY | 44.18571229 | 7.269945145 | 1590 |
| **30** | ITALY | 44.18629699 | 7.269730568 | 1590 |
| **31** | ITALY | 44.18758947 | 7.269945145 | 1590 |
| **32** | ITALY | 44.18820492 | 7.270116806 | 1590 |
| **33** | ITALY | 44.18555842 | 7.269644737 | 1590 |
| **34** | ITALY | 44.18525068 | 7.269558907 | 1590 |
| **35** | ITALY | 44.33818402 | 7.054252625 | 1250 |
| **36** | ITALY | 44.33822815 | 7.053002715 | 1300 |
| **37** | ITALY | 44.33818019 | 7.053407729 | 1300 |
| **38** | ITALY | 44.32817686 | 7.028428316 | 1460 |
| **39** | ITALY | 44.28298521 | 7.067062855 | 1300 |
| **40** | ITALY | 44.25864748 | 7.045326233 | 1675 |

| | | | | |
|---|---|---|---|---|
| **41** | ITALY | 44.258732 | 7.046248913 | 1680 |
| **42** | ITALY | 44.25848611 | 7.046324015 | 1680 |
| **43** | ITALY | 46.13935185 | 11.47203326 | 1260 |
| **44** | ITALY | 46.14764022 | 11.45871878 | 1325 |
| **45** | ITALY | 46.14686719 | 11.45973802 | 1310 |
| **46** | ITALY | 46.14688206 | 11.45939469 | 1310 |
| **47** | ITALY | 46.14697125 | 11.46019936 | 1312 |
| **48** | ITALY | 46.12890597 | 11.47687197 | 1100 |
| **49** | ITALY | 46.12956029 | 11.47652864 | 1100 |
| **50** | ITALY | 46.41701013 | 11.72039509 | 1550 |
| **51** | ITALY | 46.41709149 | 11.72057748 | 1550 |
| **52** | ITALY | 46.41730598 | 11.72089934 | 1550 |
| **53** | ITALY | 46.42127774 | 11.70846462 | 1430 |
| **54** | ITALY | 46.42148483 | 11.7081213 | 1433 |
| **55** | ITALY | 46.42175108 | 11.70782089 | 1440 |
| **56** | ITALY | 46.41805302 | 11.71061039 | 1500 |
| **57** | ITALY | 46.49079209 | 11.78062677 | 1725 |
| **58** | ITALY | 46.5021144 | 11.79169893 | 1865 |
| **59** | ITALY | 46.5012799 | 11.79201007 | 1865 |
| **60** | ITALY | 46.53169049 | 11.78120613 | 1890 |
| **61** | ITALY | 46.48254075 | 11.73736811 | 1520 |
| **62** | ITALY | 46.48289535 | 11.73702478 | 1530 |
| **63** | ITALY | 46.48596849 | 11.73048019 | 1580 |
| **64** | ITALY | 46.48309482 | 11.73720717 | 1570 |
| **65** | ITALY | 46.48255552 | 11.73642397 | 1520 |
| **66** | ITALY | 46.32211185 | 11.59942746 | 1100 |
| **67** | ITALY | 46.32314174 | 11.60028577 | 1090 |
| **68** | ITALY | 46.29718847 | 11.50802851 | 1080 |
| **69** | ITALY | 46.29781854 | 11.50733113 | 1120 |
| **70** | ITALY | 46.30080577 | 11.50885463 | 1160 |
| **71** | ITALY | 46.30090212 | 11.5105176 | 1180 |
| **72** | ITALY | 46.27468616 | 11.36267424 | 1200 |
| **73** | ITALY | 46.27468616 | 11.36267424 | 1245 |
| **74** | ITALY | 46.28023289 | 11.37005568 | 1240 |
| **75** | ITALY | 46.26668394 | 11.35178447 | 850 |
| **76** | ITALY | 46.18130125 | 11.43693924 | 1760 |
| **77** | ITALY | 46.18219266 | 11.43666029 | 1780 |
| **78** | ITALY | 46.18295034 | 11.43418193 | 1770 |
| **80** | ITALY | 46.31716958 | 11.65374756 | 1320 |
| **81** | ITALY | 46.31501321 | 11.65921926 | 1350 |
| **82** | ITALY | 46.29572072 | 11.7800796 | 1930 |
| **83** | ITALY | 46.29533525 | 11.78065896 | 1933 |
| **85** | ITALY | 46.30967752 | 11.74735665 | 1530 |
| **86** | ITALY | 46.30669819 | 11.74003959 | 1510 |
| **87** | ITALY | 46.30737263 | 11.74228191 | 1515 |

| | | | | |
|---|---|---|---|---|
| **88** | ITALY | 46.29524629 | 11.79241776 | 1940 |
| **89** | ITALY | 46.34040253 | 11.79800749 | 2015 |
| **90** | ITALY | 46.27271353 | 11.79880142 | 1610 |
| **91** | ITALY | 45.99986087 | 10.49755454 | 1496 |
| **92** | ITALY | 46.00041983 | 10.49695373 | 1496 |
| **93** | ITALY | 46.00025587 | 10.49746871 | 1496 |
| **94** | ITALY | 46.00009191 | 10.49753308 | 1496 |
| **95** | ITALY | 45.97919759 | 10.53968668 | 960 |
| **96** | ITALY | 45.97928706 | 10.53957939 | 960 |
| **97** | ITALY | 45.91366072 | 11.23493671 | 1445 |
| **98** | ITALY | 45.91449674 | 11.23356342 | 1445 |
| **99** | ITALY | 45.91503417 | 11.23253345 | 1445 |
| **100** | ITALY | 45.91509389 | 11.2319541 | 1445 |
| **101** | ITALY | 46.41478373 | 10.78331709 | 1400 |
| **102** | ITALY | 46.41173616 | 10.78938961 | 1360 |
| **103** | ITALY | 46.41030108 | 10.7920289 | 1360 |
| **104** | ITALY | 46.41049341 | 10.79022646 | 1360 |
| **105** | ITALY | 46.40720147 | 10.69576979 | 1740 |
| **106** | ITALY | 46.40885117 | 10.69134951 | 1770 |
| **107** | ITALY | 46.40998299 | 10.69067359 | 1780 |
| **108** | ITALY | 46.40975367 | 10.68962216 | 1800 |
| **109** | ITALY | 46.40888815 | 10.69173574 | 1770 |
| **110** | ITALY | 46.51437205 | 12.08912373 | 1555 |
| **111** | ITALY | 46.4951721 | 12.07562685 | 1840 |
| **113** | ITALY | 46.49438919 | 12.07446814 | 1860 |
| **114** | ITALY | 46.51455663 | 12.09148407 | 1530 |
| **115** | ITALY | 46.62431595 | 12.65882492 | 1340 |
| **116** | ITALY | 46.62226752 | 12.65378237 | 1332 |
| **117** | ITALY | 46.62263595 | 12.65377164 | 1332 |
| **118** | ITALY | 46.62316648 | 12.65355706 | 1332 |
| **119** | ITALY | 46.53545464 | 12.27400303 | 1165 |
| **120** | ITALY | 46.57585481 | 12.25215912 | 1740 |
| **122** | ITALY | 46.54652415 | 12.24321127 | 1377 |
| **123** | ITALY | 46.54609617 | 12.24175215 | 1377 |
| **124** | ITALY | 46.54627326 | 12.24100113 | 1377 |
| **127** | ITALY | 46.62030743 | 12.62799025 | 1288 |
| **128** | ITALY | 46.62024111 | 12.6282692 | 1288 |
| **129** | ITALY | 46.62016742 | 12.62835503 | 1288 |
| **130** | ITALY | 46.61071229 | 12.62099504 | 1288 |
| **131** | ITALY | 46.46859114 | 13.6611557 | 1006 |
| **132** | ITALY | 46.46969956 | 13.66158485 | 1006 |
| **133** | ITALY | 46.54581577 | 13.01414251 | 750 |
| **134** | ITALY | 46.54631754 | 13.0148077 | 750 |
| **135** | ITALY | 46.55052338 | 13.02959204 | 1300 |
| **136** | ITALY | 46.55052338 | 13.02959204 | 1300 |

| | | | | |
|---|---|---|---|---|
| **137** | ITALY | 46.55052338 | 13.02959204 | 1300 |
| **138** | ITALY | 46.55052338 | 13.02959204 | 1300 |
| **139** | ITALY | 46.55052338 | 13.02959204 | 1300 |
| **140** | ITALY | 46.55052338 | 13.02959204 | 1300 |
| **141** | ITALY | 46.55052338 | 13.02959204 | 1300 |
| **142** | ITALY | 46.51667559 | 13.67918015 | 1250 |
| **143** | ITALY | 46.51667559 | 13.67918015 | 1250 |
| **144** | ITALY | 46.51667559 | 13.67918015 | 1250 |
| **145** | ITALY | 46.51667559 | 13.67918015 | 1250 |
| **146** | ITALY | 46.51667559 | 13.67918015 | 1250 |
| **147** | ITALY | 46.51667559 | 13.67918015 | 1250 |
| **148** | ITALY | 46.46829556 | 13.66094112 | 1006 |
| **149** | ITALY | 46.4672019 | 13.66094112 | 1006 |
| **150** | ITALY | 46.46114201 | 13.6648035 | 1010 |
| **151** | ITALY | 46.46399468 | 13.66538286 | 1010 |
| **152** | ITALY | 46.43548347 | 13.56790066 | 920 |
| **153** | ITALY | 46.390872 | 13.47582579 | 1190 |
| **154** | ITALY | 46.48908578 | 11.84815407 | 1800 |
| **155** | ITALY | 46.4669876 | 12.4736774 | 1000 |
| **156** | ITALY | 46.70400477 | 10.90142012 | 1540 |
| **157** | ITALY | 46.70432114 | 10.90139866 | 1540 |
| **158** | ITALY | 46.70454922 | 10.90110898 | 1540 |
| **159** | ITALY | 46.82992077 | 11.29863381 | 1520 |
| **160** | ITALY | 46.82942163 | 11.29853725 | 1520 |
| **161** | ITALY | 46.834578 | 11.3054198 | 1750 |
| **162** | ITALY | 46.83448258 | 11.30525887 | 1750 |
| **163** | ITALY | 46.85735918 | 11.34645224 | 1780 |
| **164** | ITALY | 47.05454739 | 12.13592291 | 1630 |
| **165** | ITALY | 47.05440851 | 12.13609457 | 1630 |
| **166** | ITALY | 47.0550883 | 12.1367383 | 1630 |
| **167** | ITALY | 47.05530758 | 12.13654518 | 1630 |
| **168** | ITALY | 47.05407227 | 12.13400245 | 1630 |
| **169** | ITALY | 46.92893092 | 11.27435446 | 1430 |
| **170** | ITALY | 46.92983943 | 11.27057791 | 1490 |
| **171** | ITALY | 46.93023507 | 11.26883984 | 1510 |
| **172** | ITALY | 46.92777328 | 11.26793861 | 1600 |
| **173** | ITALY | 46.83604958 | 12.23653793 | 1420 |
| **174** | ITALY | 46.83587343 | 12.23641992 | 1425 |
| **175** | ITALY | 46.83285682 | 12.23266482 | 1500 |
| **176** | ITALY | 46.83431744 | 12.23382354 | 1500 |
| **177** | ITALY | 46.84595695 | 12.25091457 | 1470 |
| **178** | ITALY | 46.84516443 | 12.2507 | 1470 |
| **179** | ITALY | 46.84595695 | 12.25091457 | 1470 |
| **180** | ITALY | 46.74252217 | 11.44336581 | 1550 |
| **181** | ITALY | 46.74473515 | 11.44479275 | 1550 |

| | | | | |
|---|---|---|---|---|
| **182** | ITALY | 46.74573501 | 11.44535065 | 1560 |
| **183** | ITALY | 46.74263981 | 11.45134807 | 1540 |
| **184** | ITALY | 46.55052338 | 13.02959204 | 1300 |
| **185** | ITALY | 46.55052338 | 13.02959204 | 1300 |
| **186** | ITALY | 45.012488 | 6.849343 | 1665 |
| **187** | ITALY | 45.013633 | 6.848688 | 1640 |
| **188** | ITALY | 45.017881 | 6.852765 | 1580 |
| **189** | ITALY | 46.230767 | 8.296094 | 1100 |
| **190** | ITALY | 46.229528 | 8.293734 | 1150 |
| **191** | ITALY | 46.229936 | 8.29354 | 1180 |
| **192** | ITALY | 46.230351 | 8.292575 | 1210 |
| **193** | ITALY | 46.234329 | 8.289013 | 1250 |
| **194** | ITALY | 46.231687 | 8.289742 | 1220 |
| **195** | ITALY | 45.754529 | 7.704228 | 1620 |
| **196** | ITALY | 45.754529 | 7.704248 | 1610 |
| **197** | ITALY | 45.753481 | 7.710428 | 1590 |
| **198** | ITALY | 45.830967 | 6.985921 | 1590 |
| **200** | ITALY | 45.830534 | 6.987004 | 1590 |
| **201** | ITALY | 45.822504 | 6.96914 | 1430 |
| **202** | FRANCE | 45.026455 | 6.656858 | 1774 |
| **203** | FRANCE | 45.02609 | 6.656914 | 1772 |
| **204** | FRANCE | 45.02579 | 6.656929 | 1769 |
| **205** | FRANCE | 45.02423 | 6.857508 | 1755 |
| **206** | FRANCE | 44.351132 | 6.495313 | 1559 |
| **207** | FRANCE | 44.351385 | 6.495627 | 1580 |
| **208** | FRANCE | 44.351246 | 6.49657 | 1596 |
| **209** | FRANCE | 44.351434 | 6.494705 | 1580 |
| **210** | FRANCE | 44.113144 | 7.286122 | 1482 |
| **211** | FRANCE | 44.11099 | 7.298147 | 1547 |
| **212** | FRANCE | 44.111293 | 7.297478 | 1543 |
| **213** | FRANCE | 44.111687 | 7.296281 | 1534 |
| **215** | AUSTRIA | 47.280364 | 11.246395 | 900 |
| **216** | AUSTRIA | 47.280364 | 11.246395 | 900 |
| **217** | AUSTRIA | 47.191404 | 14.279823 | 1080 |
| **218** | AUSTRIA | 47.191404 | 14.279823 | 1080 |
| **219** | AUSTRIA | 47.191404 | 14.279823 | 1080 |
| **220** | AUSTRIA | 47.601707 | 13.371048 | 885 |
| **221** | AUSTRIA | 47.601707 | 13.371048 | 885 |
| **222** | AUSTRIA | 47.601707 | 13.371048 | 885 |
| **223** | AUSTRIA | 47.891831 | 14.973249 | 1000 |
| **224** | AUSTRIA | 47.891831 | 14.973249 | 1000 |
| **225** | AUSTRIA | 47.891831 | 14.973249 | 1000 |
| **226** | AUSTRIA | 47.55504 | 14.487534 | 1320 |
| **227** | AUSTRIA | 47.55504 | 14.487534 | 1320 |
| **229** | AUSTRIA | 46.659864 | 14.065762 | 840 |

| 232 | AUSTRIA | 46.899616 | 15.15976 | 1120 |
|---|---|---|---|---|
| 233 | AUSTRIA | 46.899616 | 15.15976 | 1120 |
| 234 | AUSTRIA | 46.918848 | 12.318892 | 1680 |
| 235 | AUSTRIA | 46.918848 | 12.318892 | 1680 |
| 236 | AUSTRIA | 46.918848 | 12.318892 | 1680 |
| 237 | AUSTRIA | 46.899616 | 15.15976 | 1120 |
| 238 | AUSTRIA | 47.286216 | 14.608126 | 1400 |
| 239 | AUSTRIA | 47.286216 | 14.608126 | 1400 |
| 240 | AUSTRIA | 47.286216 | 14.608126 | 1400 |
| 241 | AUSTRIA | 47.379057 | 13.573952 | 1000 |
| 242 | AUSTRIA | 47.379057 | 13.573952 | 1000 |
| 243 | AUSTRIA | 47.379057 | 13.573952 | 1000 |
| 244 | AUSTRIA | 47.764772 | 13.22402 | 700 |
| 245 | AUSTRIA | 47.764772 | 13.22402 | 700 |
| 246 | AUSTRIA | 47.764772 | 13.22402 | 700 |
| 247 | AUSTRIA | 47.091441 | 11.45359 | 1220 |
| 248 | AUSTRIA | 47.091441 | 11.45359 | 1220 |
| 249 | AUSTRIA | 47.091441 | 11.45359 | 1220 |
| 250 | AUSTRIA | 46.748477 | 14.563622 | 1040 |
| 251 | AUSTRIA | 46.748477 | 14.563622 | 1040 |
| 252 | AUSTRIA | 46.748477 | 14.563622 | 1040 |
| 253 | AUSTRIA | 47.573688 | 15.232029 | 1590 |
| 255 | AUSTRIA | 47.573688 | 15.232029 | 1590 |
| 256 | AUSTRIA | 47.565755 | 15.804348 | 1280 |
| 257 | AUSTRIA | 47.565755 | 15.804348 | 1280 |
| 258 | AUSTRIA | 47.565755 | 15.804348 | 1280 |
| 259 | AUSTRIA | 47.409089 | 16.041278 | 490 |
| 261 | AUSTRIA | 47.409089 | 16.041278 | 490 |
| 262 | AUSTRIA | 48.415644 | 15.50231 | 415 |
| 263 | AUSTRIA | 48.415644 | 15.50231 | 415 |
| 264 | AUSTRIA | 48.415644 | 15.50231 | 415 |
| 265 | AUSTRIA | 48.05046 | 15.517201 | 580 |
| 266 | AUSTRIA | 48.05046 | 15.517201 | 580 |
| 267 | AUSTRIA | 48.05046 | 15.517201 | 580 |
| 268 | AUSTRIA | 48.32664 | 15.381761 | 620 |
| 270 | AUSTRIA | 48.32664 | 15.381761 | 620 |
| 271 | AUSTRIA | 48.08883 | 13.412944 | 645 |
| 274 | AUSTRIA | 46.868997 | 14.931048 | 1025 |
| 276 | AUSTRIA | 46.868997 | 14.931048 | 1025 |
| 277 | AUSTRIA | 46.722991 | 14.659286 | 1040 |
| 279 | AUSTRIA | 46.722991 | 14.659286 | 1040 |
| 280 | AUSTRIA | 47.588148 | 15.742625 | 960 |
| 281 | AUSTRIA | 47.588148 | 15.742625 | 960 |
| 282 | AUSTRIA | 47.588148 | 15.742625 | 960 |
| 283 | AUSTRIA | 48.154327 | 16.109937 | 405 |

| | | | | |
|---|---|---|---|---|
| **284** | AUSTRIA | 48.154327 | 16.109937 | 405 |
| **285** | AUSTRIA | 48.154327 | 16.109937 | 405 |
| **286** | AUSTRIA | 47.071393 | 10.633564 | 1505 |
| **287** | AUSTRIA | 47.071393 | 10.633564 | 1505 |
| **288** | AUSTRIA | 47.071393 | 10.633564 | 1505 |
| **290** | AUSTRIA | 47.182552 | 12.266278 | 1680 |
| **291** | AUSTRIA | 47.182552 | 12.266278 | 1680 |
| **292** | AUSTRIA | 47.373201 | 10.685449 | 1450 |
| **293** | AUSTRIA | 47.373201 | 10.685449 | 1450 |
| **294** | AUSTRIA | 47.373201 | 10.685449 | 1450 |
| **295** | AUSTRIA | 47.430257 | 12.354877 | 1360 |
| **300** | AUSTRIA | 47.438273 | 15.115433 | 1290 |
| **304** | AUSTRIA | 47.605121 | 12.103071 | 1040 |
| **307** | AUSTRIA | 47.846391 | 14.845474 | 1325 |
| **310** | AUSTRIA | 48.097673 | 13.119575 | 465 |
| **315** | AUSTRIA | 46.498946 | 14.643016 | 1080 |
| **320** | AUSTRIA | 47.220326 | 15.785514 | 560 |
| **325** | AUSTRIA | 48.375031 | 15.383445 | 640 |
| **330** | AUSTRIA | 46.74919 | 15.154631 | 1065 |
| **335** | AUSTRIA | 48.319402 | 15.027712 | 1000 |
| **339** | AUSTRIA | 46.587064 | 13.776855 | 1120 |
| **340** | SWITZERLAND | 47.64560829 | 8.74410406 | 450 |
| **341** | SWITZERLAND | 47.64560829 | 8.74410406 | 450 |
| **342** | SWITZERLAND | 47.64560829 | 8.74410406 | 450 |
| **344** | SWITZERLAND | 47.63568582 | 9.13815627 | 530 |
| **345** | SWITZERLAND | 47.63568582 | 9.13815627 | 530 |
| **346** | SWITZERLAND | 47.45708745 | 9.32027616 | 903 |
| **347** | SWITZERLAND | 47.45708745 | 9.32027616 | 903 |
| **348** | SWITZERLAND | 47.45708745 | 9.32027616 | 903 |
| **349** | SWITZERLAND | 47.20682328 | 7.03419565 | 1035 |
| **350** | SWITZERLAND | 47.20682328 | 7.03419565 | 1035 |
| **351** | SWITZERLAND | 47.20682328 | 7.03419565 | 1035 |
| **352** | SWITZERLAND | 47.31632067 | 7.40269057 | 655 |
| **353** | SWITZERLAND | 47.31632067 | 7.40269057 | 655 |
| **354** | SWITZERLAND | 47.31632067 | 7.40269057 | 655 |
| **355** | SWITZERLAND | 47.22661252 | 7.83178805 | 480 |
| **357** | SWITZERLAND | 47.22661252 | 7.83178805 | 480 |
| **358** | SWITZERLAND | 47.36547457 | 7.82263008 | 865 |
| **359** | SWITZERLAND | 47.36547457 | 7.82263008 | 865 |
| **360** | SWITZERLAND | 47.36547457 | 7.82263008 | 865 |
| **361** | SWITZERLAND | 47.16551536 | 9.28739407 | 1640 |
| **362** | SWITZERLAND | 47.16551536 | 9.28739407 | 1640 |
| **363** | SWITZERLAND | 47.26273861 | 9.42940434 | 1460 |
| **364** | SWITZERLAND | 47.26273861 | 9.42940434 | 1460 |
| **365** | SWITZERLAND | 47.26273861 | 9.42940434 | 1460 |

| | | | | |
|---|---|---|---|---|
| **366** | SWITZERLAND | 47.04378493 | 7.59787919 | 620 |
| **367** | SWITZERLAND | 47.04378493 | 7.59787919 | 620 |
| **368** | SWITZERLAND | 47.04378493 | 7.59787919 | 620 |
| **369** | SWITZERLAND | 47.02175967 | 7.85270326 | 1165 |
| **370** | SWITZERLAND | 47.02175967 | 7.85270326 | 1165 |
| **371** | SWITZERLAND | 47.02175967 | 7.85270326 | 1165 |
| **372** | SWITZERLAND | 46.98249309 | 8.84995893 | 1490 |
| **374** | SWITZERLAND | 46.98249309 | 8.84995893 | 1490 |
| **376** | SWITZERLAND | 47.04802314 | 8.7131424 | 1180 |
| **378** | SWITZERLAND | 47.09604765 | 9.1873817 | 1120 |
| **380** | SWITZERLAND | 47.09604765 | 9.1873817 | 1120 |
| **381** | SWITZERLAND | 47.1093514 | 9.02649182 | 1460 |
| **382** | SWITZERLAND | 47.1093514 | 9.02649182 | 1460 |
| **385** | SWITZERLAND | 47.07806309 | 9.25058022 | 1700 |
| **386** | SWITZERLAND | 46.97507113 | 9.78725991 | 1540 |
| **388** | SWITZERLAND | 46.75785885 | 6.46841406 | 1130 |
| **392** | SWITZERLAND | 46.79710185 | 7.00496603 | 610 |
| **394** | SWITZERLAND | 46.79710185 | 7.00496603 | 610 |
| **395** | SWITZERLAND | 46.8423952 | 8.00545824 | 1170 |
| **396** | SWITZERLAND | 46.8423952 | 8.00545824 | 1170 |
| **398** | SWITZERLAND | 46.77391207 | 8.33660654 | 1740 |
| **400** | SWITZERLAND | 46.90255281 | 8.32981578 | 1500 |
| **401** | SWITZERLAND | 46.90255281 | 8.32981578 | 1500 |
| **402** | SWITZERLAND | 46.77469683 | 8.73527876 | 1460 |
| **404** | SWITZERLAND | 46.77469683 | 8.73527876 | 1460 |
| **405** | SWITZERLAND | 46.79100679 | 9.04987194 | 1520 |
| **406** | SWITZERLAND | 46.79100679 | 9.04987194 | 1520 |
| **408** | SWITZERLAND | 46.75285117 | 9.39301711 | 1500 |
| **410** | SWITZERLAND | 46.75285117 | 9.39301711 | 1500 |
| **411** | SWITZERLAND | 46.90924229 | 9.42832499 | 1020 |
| **412** | SWITZERLAND | 46.90924229 | 9.42832499 | 1020 |
| **414** | SWITZERLAND | 46.87704426 | 9.80245449 | 1740 |
| **416** | SWITZERLAND | 46.59078608 | 6.16219861 | 1265 |
| **418** | SWITZERLAND | 46.65812784 | 7.16304745 | 1270 |
| **420** | SWITZERLAND | 46.65812784 | 7.16304745 | 1270 |
| **421** | SWITZERLAND | 46.51003045 | 7.54756334 | 1550 |
| **422** | SWITZERLAND | 46.51003045 | 7.54756334 | 1550 |
| **424** | SWITZERLAND | 46.699867 | 7.76395538 | 1520 |
| **427** | SWITZERLAND | 46.6582495 | 8.53320812 | 1360 |
| **428** | SWITZERLAND | 46.6582495 | 8.53320812 | 1360 |
| **430** | SWITZERLAND | 46.53274412 | 8.36579642 | 1670 |
| **431** | SWITZERLAND | 46.53274412 | 8.36579642 | 1670 |
| **432** | SWITZERLAND | 46.53274412 | 8.36579642 | 1670 |
| **434** | SWITZERLAND | 46.58437271 | 9.13285987 | 1725 |
| **436** | SWITZERLAND | 46.70439244 | 8.92097277 | 1640 |

| 438 | SWITZERLAND | 46.70439244 | 8.92097277 | 1640 |
|---|---|---|---|---|
| 439 | SWITZERLAND | 46.70479772 | 9.19115549 | 960 |
| 440 | SWITZERLAND | 46.70479772 | 9.19115549 | 960 |
| 442 | SWITZERLAND | 46.54413293 | 9.31789976 | 1630 |
| 444 | SWITZERLAND | 46.54413293 | 9.31789976 | 1630 |
| 445 | SWITZERLAND | 46.66836787 | 10.06171181 | 1670 |
| 446 | SWITZERLAND | 46.66836787 | 10.06171181 | 1670 |
| 448 | SWITZERLAND | 46.29332706 | 7.23490265 | 1630 |
| 450 | SWITZERLAND | 46.4489104 | 6.918335 | 920 |
| 451 | SWITZERLAND | 46.4489104 | 6.918335 | 920 |
| 453 | SWITZERLAND | 46.35868883 | 6.97130339 | 1225 |
| 454 | SWITZERLAND | 46.35868883 | 6.97130339 | 1225 |
| 456 | SWITZERLAND | 46.45078563 | 7.51350254 | 1850 |
| 458 | SWITZERLAND | 46.33079901 | 7.54058404 | 1490 |
| 460 | SWITZERLAND | 46.33079901 | 7.54058404 | 1490 |
| 461 | SWITZERLAND | 46.39823609 | 7.78450259 | 1600 |
| 463 | SWITZERLAND | 46.39823609 | 7.78450259 | 1600 |
| 465 | SWITZERLAND | 46.36536073 | 8.17153512 | 1680 |
| 467 | SWITZERLAND | 46.38453564 | 8.12629561 | 1250 |
| 469 | SWITZERLAND | 46.38453564 | 8.12629561 | 1250 |
| 470 | SWITZERLAND | 46.38199651 | 8.01707381 | 2010 |
| 471 | SWITZERLAND | 46.38199651 | 8.01707381 | 2010 |
| 473 | SWITZERLAND | 46.27886891 | 8.47579076 | 1380 |
| 475 | SWITZERLAND | 46.31857028 | 8.52095467 | 1280 |
| 477 | SWITZERLAND | 46.45532806 | 9.19264644 | 1620 |
| 479 | SWITZERLAND | 46.45532806 | 9.19264644 | 1620 |
| 480 | SWITZERLAND | 46.27404204 | 9.14834001 | 1050 |
| 482 | SWITZERLAND | 46.27404204 | 9.14834001 | 1050 |
| 484 | SWITZERLAND | 46.32439664 | 9.55064465 | 1590 |
| 486 | SWITZERLAND | 46.297474 | 10.05252477 | 1430 |
| 488 | SWITZERLAND | 46.14846407 | 7.10438891 | 905 |
| 489 | SWITZERLAND | 46.14846407 | 7.10438891 | 905 |
| 490 | SWITZERLAND | 46.14846407 | 7.10438891 | 905 |
| 493 | SWITZERLAND | 46.22334638 | 7.44512528 | 1630 |
| 495 | SWITZERLAND | 46.25937612 | 7.51440954 | 780 |
| 497 | SWITZERLAND | 46.25937612 | 7.51440954 | 780 |
| 498 | SWITZERLAND | 46.21572485 | 7.46286487 | 1970 |
| 499 | SWITZERLAND | 46.21572485 | 7.46286487 | 1970 |

Supplement 2: List of SNPs submitted to 384 Illumina GoldenGate genotyping assay. Total 94 SNPs originated from CRSP dataset, 102 SNPs from Uppsala (Sweden) and 188 SNPs from Arborea dataset (Canada)

## CRSP dataset

| SNP name | Source | Annotation |
|---|---|---|
| 2_9603_01-Paab_139 | CRSP | NA |
| 2_4196_01-Paab_201 | CRSP | GTP binding protein, Similar to *Arabidopsis thaliana* (NP_569023), e=3e-17 & max.identity=35% |
| UMN_853_01-Paab_38 | CRSP | Anonymous locus |
| UMN_3055_01-Paab_224 | CRSP | Similar to Arabidopsis thaliana (NP_177203), e=1e-48 & max.identity=78%, protein root hair specific 10 |
| CL1694Contig1_02-365 | CRSP | Similar to Arabidopsis thaliana (NP_172112), e=1e-99 & max.identity=84%, U5 small nuclear ribonucleoprotein component, 116 kD, GTPase activity, GTP binding, translation factor activity, nucleic acid binding,translation elongation factor activity, INVOLVED IN: regulation of embryo sac egg cell differentiation, embryo development ending in seed dormancy |
| 2_9665_01-Paab_175 | CRSP | Similar to Ricinus communis (XP_002509420.1), e=5e-86 &max.identity=62%, interferon-induced guanylate-binding protein, putative |
| CL717Contig1_05-Paab_95 | CRSP | Similar to Medicago truncatula (XP_003601864.1), e=1e-20 & max. identity=61%, photosystem II core complex proteins psbY, chloroplast precursor |
| 0_13978_01-Paab_102 | CRSP | binding protein |
| 0_177_01-Paab_165 | CRSP | Anonymous locus |
| 0_8531_01-Paab_157 | CRSP | Similar to Ricinus communis (XP_002518462.1), e=2e-55 & max.identity=62%, multicopper oxidase, putative |
| CL3507Contig1_03-Paab_191 | CRSP | Similar to Arabidopsis thaliana, (NP_566474), e=3e-82 &max.identity=65%, Ca2+ antiporter/cation exchanger, cation/calcium exchanger 3 |
| 0_13957_02-Paab_309 | CRSP | receptor-like protein kinase HSL1-like (leucine-rich repeat receptor-like protein kinase) |
| 2_7725_01-Paab_466 | CRSP | Similar to Arabidopsis thaliana (NP_001189624.1), e=4e-85 & max. identity=62%, beta-galactosidase 8 |
| 0_13957_02-132 | CRSP | receptor-like protein kinase HSL1-like (leucine-rich repeat receptor-like protein kinase), Anonymous locus |
| CL3771Contig1_04-Paab_419 | CRSP | Ubiquitin carrier protein E, ubiquitin-conjugating enzyme E2 32-like |
| CL3582Contig1_03-Paab_63 | CRSP | trehalose-6-phosphate synthase (TPS) |

| | | |
|---|---|---|
| 0_15639_01-392 | CRSP | Anonymous locus |
| 0_13680_01-Paab_216 | CRSP | VAC14-like protein |
| 0_7171_01-Paab_233 | CRSP | Glycogen synthase kinase-3 beta, putative, Anonymous locus |
| 0_9457_01-Paab_421 | CRSP | pentatricopeptide repeat-containing protein, putative |
| CL304Contig1_01-202 | CRSP | Oxygen-evolving enhancer protein 1, chloroplast precursor, putative |
| CL1758Contig1_04-288 | CRSP | NADH dehydrogenase [ubiquinone] iron-sulfur protein 8,mitochondrial electron transport |
| 2_9280_01-Paab_123 | CRSP | Anonymous locus, chromatin remodeling complex subunit |
| 0_11772_01-Paab_103 | CRSP | o-sialoglycoprotein endopeptidase |
| 2_9280_01-Paab_338 | CRSP | Anonymous locus, chromatin remodeling complex subunit |
| 0_8531_01-Paab_363 | CRSP | multicopper oxidase, putative |
| CL1692Contig1_05-Paab_178 | CRSP | histone ubiquitination proteins group, chromosome segregation protein SMC, primarily archaeal type |
| 0_4541_02-266 | CRSP | BTB/POZ domain-containing protein |
| 2_9280_01-Paab_193 | CRSP | chromatin remodeling complex subunit |
| CL304Contig1_01-Paab_118 | CRSP | Oxygen-evolving enhancer protein 1, chloroplast precursor, putative |
| 2_10438_01-351 | CRSP | amidophosphoribosyltransferase, putative |
| 0_9383_01-Paab_438 | CRSP | ubiquitin carboxyl-terminal hydrolase-like protein |
| 0_366_02-Paab_380 | CRSP | Heat stress transcription factor B-4 |
| 2_4892_01-Paab_39 | CRSP | ccaat-binding transcription factor, putative |
| 2_3947_01-Paab_298 | CRSP | AT hook motif DNA-binding family protein |
| 2_9845_01-Paab_282 | CRSP | Papain family cysteine protease |
| CL4284Contig1_01-Paab_180 | CRSP | Anonymous locus |
| 0_1439_01-Paab_226 | CRSP | Pre-mRNA-splicing factor 38B |
| PabiesCol1_583 | CRSP | picea abies col1 gene for constans-like 1 |
| 0_17587_01-Paab_42 | CRSP | Anonymous locus |
| 2_6491_01-Paab_360 | CRSP | Coiled-coil domain-containing protein, putative |
| 0_9457_01-Paab_46 | CRSP | pentatricopeptide repeat-containing protein, putative |
| 0_10754_01-Paab_320 | CRSP | FACT complex subunit, global transcription factor group, partial |
| 2_5636_01-Paab_399 | CRSP | pentatricopeptide repeat-containing protein |
| 0_9749_01-Paab_337 | CRSP | Serine/threonine-protein kinase PBS1, putative |
| 0_3128_02-Paab_79 | CRSP | f-box family protein |
| UMN_1023_01-267 | CRSP | F-box/LRR-repeat protein 14 |
| 0_2433_01-Paab_290 | CRSP | histidine triad family protein |
| 2_4976_01-Paab_176 | CRSP | 26S proteasome regulatory subunit N6 |
| 0_7171_01-Paab_359 | CRSP | Glycogen synthase kinase-3 beta, putative |
| CL4511Contig1_02-Paab_223 | CRSP | oligopeptidase, putative [Ricinus communis) calmodulin-binding region domain containing protein (Oryza sativa) |
| 2_5636_01-Paab_209 | CRSP | pentatricopeptide repeat-containing protein |
| UMN_4748_01-Paab_38 | CRSP | peptide transporter |

| | | |
|---|---|---|
| 2_9087_01-Paab_39 | CRSP | Anonymous locus |
| 0_10267_01-Paab_148 | CRSP | myb domain protein 55 |
| UMN_2809_01-104 | CRSP | Hypothetical protein |
| 0_10515_01-Paab_158 | CRSP | Anonymous locus |
| CL1308Contig1_03-Paab_181 | CRSP | NA |
| CL1148Contig1_08-Paab_134 | CRSP | malate dehydrogenase |
| UMN_4091_02-Paab_137 | CRSP | F box family protein |
| 2_4586_01-Paab_365 | CRSP | Oligosaccharyltransferase complex/magnesium transporter family protein |
| 2_9328_01-Paab_425 | CRSP | transducin/WD40 domain-containing protein |
| CL1694Contig1_01-Paab_235 | CRSP | U5 small nuclear ribonucleoprotein component, 116 kD |
| 2_2937_01-127 | CRSP | ATP/ADP transporter, partial |
| CL1343Contig1_05-Paab_165 | CRSP | Phosphoenolpyruvate carboxykinase [ATP], putative |
| 0_17215_01-Paab_225 | CRSP | Magnesium-chelatase subunit H, putative |
| CL3602Contig1_03-Paab_219 | CRSP | protochlorophyllide reductase B, |
| CL1148Contig1_08-Paab_225 | CRSP | malate dehydrogenase |
| 2_3851_01-Paab_280 | CRSP | Anonymous locus |
| UMN_4091_02-Paab_39 | CRSP | F box family protein |
| 0_10631_01-Paab_193 | CRSP | heat-shock protein 70T-2 |
| CL3795Contig1_01-Paab_45 | CRSP | amino acid dehydrogenase family protein |
| CL1694Contig1_04-Paab_90 | CRSP | U5 small nuclear ribonucleoprotein component, 116 kD |
| 0_2354_01-Paab_194 | CRSP | armadillo repeat only 1 protein |
| 2_7803_01-Paab_235 | CRSP | glycoside hydrolase family 28 protein / polygalacturonase (pectinase) family protein |
| 2_3867_02-Paab_440 | CRSP | profilin, putative |
| 2_4723_01-Paab_276 | CRSP | coatomer gamma subunit, putative |
| 0_489_01-Paab_316 | CRSP | Anonymous locus |
| 0_12021_01-161 | CRSP | STRUBBELIG-receptor family 6-protein serine/threonine kinase activity, protein kinase activity, ATP binding |
| 2_9466_01-Paab_179 | CRSP | Protease ecfE, putative (present in plastid and chloroplast) |
| 0_7471_01-Paab_399 | CRSP | Similar to EX309936.1 [Picea glauca] |
| 0_7921_01-Paab_212 | CRSP | short chain dehydrogenase, putative |
| 0_13058_01-Paab_551 | CRSP | Polygalacturonase |
| 0_9457_01-Paab_115 | CRSP | pentatricopeptide repeat-containing protein, putative |
| UMN_1604_01-Paab_348 | CRSP | SNF2 family DNA-dependent ATPase |
| 0_17215_01-Paab_108 | CRSP | Magnesium-chelatase subunit H, putative |
| 0_14976_01-Paab_305 | CRSP | ATP binding protein, putative |
| 0_11090_01-Paab_251 | CRSP | protein binding protein, putative |
| UMN_7021_02-141 | CRSP | Hypotetical protein |
| 0_8111_01-39 | CRSP | 3-hydroxyisobutyrate dehydrogenase |
| CL3771Contig1_04-Paab_68 | CRSP | Ubiquitin carrier protein E, ubiquitin-conjugating enzyme E2 32-like |

| | | |
|---|---|---|
| 0_13957_02-Paab_27 | CRSP | receptor-like protein kinase HSL1-like (leucine-rich repeat receptor-like protein kinase) |
| CL4257Contig1_01-Paab_391 | CRSP | Hypotetical protein |
| CL1530Contig1_04-Paab_64 | CRSP | histone H2B - performs performs essential roles in maintaining structural integrity of the nucleosome, chromatin condensation, and binding |

**Arborea dataset (Canada)**

| SNP name | Source | Annotation |
|---|---|---|
| PabiesPRR1_1632 | Canada | Picea abies Putative PRR1 gene |
| PabiesZTL_397 | Canada | Picea likiangensis isolate PLJT-TBS-2 putative ZTL (ztl) gene, partial cds |
| PaPHYN_RI204 | Canada | Picea abies partial phynrI gene for phytochrome N |
| PaPHYO_RI145 | Canada | Picea abies partial phyo gene for phytochrome O |
| PGLM2-0013 | Canada | Anonymous locus |
| PGLM2-0021 | Canada | NA |
| PGLM2-0024 | Canada | KH domain-containing protein |
| PGLM2-0049 | Canada | oxidoreductase, 2OG-Fe(II) oxygenase family protein |
| PGLM2-0081 | Canada | galactosyltransferase family protein |
| PGLM2-0098 | Canada | PP2AA2 (PROTEIN PHOSPHATASE 2A SUBUNIT A2); protein phosphatase type 2A regulator |
| PGLM2-0127 | Canada | PHS2 (ALPHA-GLUCAN PHOSPHORYLASE 2); phosphorylase/ transferase, transferring glycosyl groups |
| PGLM2-0130 | Canada | Clone WS02756_M22 unknown mRNA |
| PGLM2-0134 | Canada | Anonymous locus |
| PGLM2-0138 | Canada | ZAC; ARF GTPase activator/ phospholipid binding |
| PGLM2-0140 | Canada | Transcribed locus, moderately similar to XP_001771189.1 predicted protein [Physcomitrella patens subsp. patens] |
| PGLM2-0154 | Canada | unknown protein |
| PGLM2-0158 | Canada | SK13 (SHAGGY-LIKE KINASE 13); ATP binding / protein kinase/ protein serine/threonine kinase |
| PGLM2-0160 | Canada | NA |
| PGLM2-0169 | Canada | ATCUL1 (ARABIDOPSIS THALIANA CULLIN 1); protein binding |
| PGLM2-0193 | Canada | LINC1 (LITTLE NUCLEI1) |
| PGLM2-0195 | Canada | LOCATED IN: endomembrane system; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: LMBR1-like conserved region (InterPro:IPR006876); BEST Arabidopsis thaliana protein match is: LMBR1 integral membrane family protein (TAIR:AT5G01460.1); Has 250 Blast hits to 249 proteins in 81 species: Archae - 0; Bacteria - 0; Metazoa - 107; Fungi - 39; Plants - 31; Viruses - 0; Other Eukaryotes - 73 (source: NCBI BLink). |
| PGLM2-0227 | Canada | Transcribed locus |
| PGLM2-0242 | Canada | SEC22; transporter |
| PGLM2-0271 | Canada | cpHsc70-1 (chloroplast heat shock protein 70-1); ATP binding |
| PGLM2-0285 | Canada | MRH1 (morphogenesis of root hair 1); ATP binding / protein binding / protein kinase/ protein serine/threonine kinase/ protein tyrosine kinase |

| PGLM2-0288 | Canada | Transcribed locus |
|---|---|---|
| PGLM2-0295 | Canada | binding |
| PGLM2-0296 | Canada | ARFA1D; GTP binding / phospholipase activator/ protein binding |
| PGLM2-0314 | Canada | ATCS; ATP binding / ATP citrate synthase/ citrate (SI)-synthase |
| PGLM2-0349 | Canada | Transcribed locus |
| PGLM2-0353 | Canada | haloacid dehalogenase-like hydrolase family protein |
| PGLM2-0360 | Canada | FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 7 growth stages; BEST Arabidopsis thaliana protein match is: proline-rich family protein (TAIR:AT3G09000.1); Has 94255 Blast hits to 49644 proteins in 1573 species: Archae - 225; Bacteria - 11215; Metazoa - 37735; Fungi - 21320; Plants - 3339; Viruses - 2662; Other Eukaryotes - 17759 (source: NCBI BLink). |
| PGLM2-0368 | Canada | FLA7 (FASCICLIN-LIKE ARABINOOGALACTAN 7) |
| PGLM2-0383 | Canada | Transcribed locus, moderately similar to XP_002283272.1 PREDICTED: hypothetical protein [Vitis vinifera] |
| PGLM2-0391 | Canada | Transcribed locus |
| PGLM2-0395 | Canada | NA |
| PGLM2-0404 | Canada | 40S ribosomal protein S15 (RPS15D) |
| PGLM2-0425 | Canada | ATCOAD (4-phosphopantetheine adenylyltransferase); nucleotidyltransferase/ pantetheine-phosphate adenylyltransferase |
| PGLM2-0433 | Canada | splicing factor Prp18 family protein |
| PGLM2-0437 | Canada | folic acid binding / transferase |
| PGLM2-0440 | Canada | Os11g0167800 |
| PGLM2-0450 | Canada | unknown protein |
| PGLM2-0460 | Canada | NLI interacting factor (NIF) family protein |
| PGLM2-0465 | Canada | unknown protein |
| PGLM2-0489 | Canada | NHL repeat-containing protein |
| PGLM2-0520 | Canada | BIN4 (brassinosteroid-insensitive4); double-stranded DNA binding |
| PGLM2-0565 | Canada | SNAP33 (SOLUBLE N-ETHYLMALEIMIDE-SENSITIVE FACTOR ADAPTOR PROTEIN 33); SNAP receptor/ protein binding |
| PGLM2-0571 | Canada | LIN2 (LESION INITIATION 2); coproporphyrinogen oxidase |
| PGLM2-0584 | Canada | thylakoid lumenal 17.4 kDa protein, chloroplast |
| PGLM2-0592 | Canada | TRN1 (TORNADO 1) |
| PGLM2-0610 | Canada | PRMT11 (ARGININE METHYLTRANSFERASE 11); protein-arginine N-methyltransferase |
| PGLM2-0624 | Canada | SDRB (SHORT-CHAIN DEHYDROGENASE-REDUCTASE B); binding / catalytic/ oxidoreductase |
| PGLM2-0642 | Canada | INVOLVED IN: biological_process unknown |
| PGLM2-0645 | Canada | bile acid:sodium symporter family protein |
| PGLM2-0674 | Canada | NIC1 (NICOTINAMIDASE 1); catalytic/ nicotinamidase |
| PGLM2-0702 | Canada | ATP binding / aminoacyl-tRNA ligase/ leucine-tRNA ligase/ nucleotide binding |
| PGLM2-0703 | Canada | mitochondrial substrate carrier family protein |
| PGLM2-0733 | Canada | NF-YA7 (NUCLEAR FACTOR Y, SUBUNIT A7); specific transcriptional repressor/ transcription factor |

| | | |
|---|---|---|
| PGLM2-0762 | Canada | Anonymous locus, unknown protein |
| PGLM2-0769 | Canada | tetratricopeptide repeat (TPR)-containing protein |
| PGLM2-0780 | Canada | PPX2 (PROTEIN PHOSPHATASE X 2); protein serine/threonine phosphatase |
| PGLM2-0784 | Canada | thioesterase family protein |
| PGLM2-0796 | Canada | Transcribed locus, moderately similar to XP_001759251.1 predicted protein [Physcomitrella patens subsp. patens] |
| PGLM2-0798 | Canada | DNAJ heat shock N-terminal domain-containing protein |
| PGLM2-0810 | Canada | WIN1 (HOPW1-1-INTERACTING 1); N2-acetyl-L-ornithine:2-oxoglutarate 5-aminotransferase/ catalytic/ pyridoxal phosphate binding / transaminase |
| PGLM2-0818 | Canada | DNAJ heat shock N-terminal domain-containing protein |
| PGLM2-0828 | Canada | CLPB4 (CASEIN LYTIC PROTEINASE B4); ATP binding / ATPase/ nucleoside-triphosphatase/ nucleotide binding / protein binding |
| PGLM2-0886 | Canada | thylakoid lumenal 20 kDa protein |
| PGLM2-0887 | Canada | AMK2 (Adenosine monophosphate kinase); ATP binding / adenylate kinase/ nucleobase, nucleoside, nucleotide kinase/ nucleotide kinase/ phosphotransferase, phosphate group as acceptor |
| PGLM2-0901 | Canada | Anonymous locus |
| PGLM2-0923 | Canada | dormancy/auxin associated family protein |
| PGLM2-0924 | Canada | DFR (DIHYDROFLAVONOL 4-REDUCTASE); dihydrokaempferol 4-reductase |
| PGLM2-0944 | Canada | oxidoreductase NAD-binding domain-containing protein |
| PGLM2-1030 | Canada | postsynaptic protein-related |
| PGLM2-1031 | Canada | Transcribed locus |
| PGLM2-1037 | Canada | COBL1 (COBRA-LIKE PROTEIN 1 PRECURSOR) |
| PGLM2-1069 | Canada | Transcribed locus |
| PGLM2-1091 | Canada | fringe-related protein |
| PGLM2-1099 | Canada | unknown protein |
| PGLM2-1147 | Canada | unknown protein |
| PGLM2-1166 | Canada | AIR9; protein binding |
| PGLM2-1170 | Canada | PEX11D |
| PGLM2-1171 | Canada | flavodoxin family protein / radical SAM domain-containing protein |
| PGLM2-1178 | Canada | FUNCTIONS IN: molecular_function unknown |
| PGLM2-1182 | Canada | SYNC1; ATP binding / aminoacyl-tRNA ligase/ asparagine-tRNA ligase/ aspartate-tRNA ligase/ nucleic acid binding / nucleotide binding |
| PGLM2-1191 | Canada | unknown protein |
| PGLM2-1223 | Canada | FUNCTIONS IN: molecular_function unknown |
| PGLM2-1269 | Canada | Anonymous locus |
| PGLM2-1273 | Canada | amino acid transporter family protein |
| PGLM2-1385 | Canada | fructose-1,6-bisphosphatase, putative / D-fructose-1,6-bisphosphate 1-phosphohydrolase, putative / FBPase, putative |
| PGLM2-1476 | Canada | aconitase C-terminal domain-containing protein |
| PGLM2-1477 | Canada | signal peptide peptidase family protein |
| PGLM2-1492 | Canada | RGLG2 (RING domain Ligase2); ubiquitin-protein ligase |
| PGLM2-1514 | Canada | Transcribed locus |
| PGLM2-1528 | Canada | SYP124 (SYNTAXIN OF PLANTS 124); SNAP receptor |

| | | |
|---|---|---|
| PGWD1-0007 | Canada | ATEBP (ETHYLENE-RESPONSIVE ELEMENT BINDING PROTEIN); DNA binding / protein binding / transcription activator/ transcription factor |
| PGWD1-0024 | Canada | PMEPCRF (PECTIN METHYLESTERASE PCR FRAGMENT F); pectinesterase |
| PGWD1-0029 | Canada | pectinacetylesterase family protein |
| PGWD1-0041 | Canada | NA |
| PGWD1-0042 | Canada | POM1 (POM-POM1); chitinase |
| PGWD1-0051 | Canada | reversibly glycosylated polypeptide, putative |
| PGWD1-0053 | Canada | ATOMT1 (O-METHYLTRANSFERASE 1); caffeate O-methyltransferase/ myricetin 3'-O-methyltransferase/ quercetin 3-O-methyltransferase |
| PGWD1-0117 | Canada | HSP18.2 (heat shock protein 18.2) |
| PGWD1-0118 | Canada | anac028 (Arabidopsis NAC domain containing protein 28); transcription factor |
| PGWD1-0132 | Canada | AtMYB93 (myb domain protein 93); DNA binding / transcription factor |
| PGWD1-0142 | Canada | Transcribed locus |
| PGWD1-0147 | Canada | ATGH9A1 (ARABIDOPSIS THALIANA GLYCOSYL HYDROLASE 9A1); cellulase/ hydrolase, hydrolyzing O-glycosyl compounds |
| PGWD1-0152 | Canada | unknown protein |
| PGWD1-0158 | Canada | CAD9 (CINNAMYL ALCOHOL DEHYDROGENASE 9); binding / catalytic/ oxidoreductase/ zinc ion binding |
| PGWD1-0184 | Canada | ATOMT1 (O-METHYLTRANSFERASE 1); caffeate O-methyltransferase/ myricetin 3'-O-methyltransferase/ quercetin 3-O-methyltransferase |
| PGWD1-0191 | Canada | Anonymous locus |
| PGWD1-0220 | Canada | PIP2A (PLASMA MEMBRANE INTRINSIC PROTEIN 2A); water channel |
| PGWD1-0242 | Canada | gibberellin-responsive protein, putative |
| PGWD1-0257 | Canada | ATFD3 (ferredoxin 3); 2 iron, 2 sulfur cluster binding / electron carrier/ iron-sulfur cluster binding |
| PGWD1-0269 | Canada | GMD1 (GDP-D-MANNOSE 4,6-DEHYDRATASE 1); GDP-mannose 4,6-dehydratase/ binding / catalytic/ coenzyme binding |
| PGWD1-0282 | Canada | GATL6; polygalacturonate 4-alpha-galacturonosyltransferase/ transferase, transferring glycosyl groups / transferase, transferring hexosyl groups |
| PGWD1-0337 | Canada | UDP-glucose 6-dehydrogenase, putative |
| PGWD1-0344 | Canada | AtGH9A4 (Arabidopsis thaliana Glycosyl Hydrolase 9A4); catalytic/ hydrolase, hydrolyzing O-glycosyl compounds |
| PGWD1-0362 | Canada | CESA1 (CELLULOSE SYNTHASE 1); cellulose synthase/ transferase, transferring glycosyl groups |
| PGWD1-0375 | Canada | NA |
| PGWD1-0391 | Canada | GAMMA-TIP (GAMMA TONOPLAST INTRINSIC PROTEIN); water channel |
| PGWD1-0396 | Canada | pectinesterase family protein |
| PGWD1-0400 | Canada | MYB31 (MYB DOMAIN PROTEIN 31); DNA binding / transcription factor |
| PGWD1-0413 | Canada | universal stress protein (USP) family protein |

| | | |
|---|---|---|
| PGWD1-0418 | Canada | xyloglucan:xyloglucosyl transferase, putative / xyloglucan endotransglycosylase, putative / endo-xyloglucan transferase, putative |
| PGWD1-0421 | Canada | LHCB4.3 (light harvesting complex PSII); chlorophyll binding |
| PGWD1-0453 | Canada | LDOX (LEUCOANTHOCYANIDIN DIOXYGENASE); leucocyanidin oxygenase |
| PGWD1-0468 | Canada | MTHFR2 (METHYLENETETRAHYDROFOLATE REDUCTASE 2); methylenetetrahydrofolate reductase (NADPH) |
| PGWD1-0485 | Canada | MYB4; DNA binding / transcription factor |
| PGWD1-0510 | Canada | unknown protein |
| PGWD1-0511 | Canada | XT2 (UDP-XYLOSYLTRANSFERASE 2); UDP-xylosyltransferase/ transferase/ transferase, transferring glycosyl groups / xyloglucan 6-xylosyltransferase |
| PGWD1-0519 | Canada | CESA1 (CELLULOSE SYNTHASE 1); cellulose synthase/ transferase, transferring glycosyl groups |
| PGWD1-0533 | Canada | WRKY65; transcription factor |
| PGWD1-0551 | Canada | unknown protein |
| PGWD1-0556 | Canada | cysteine proteinase, putative |
| PGWD1-0557 | Canada | ribulose bisphosphate carboxylase small chain 1B / RuBisCO small subunit 1B (RBCS-1B) (ATS1B) |
| PGWD1-0578 | Canada | TUB6 (BETA-6 TUBULIN); structural constituent of cytoskeleton |
| PGWD1-0586 | Canada | PIP2;8 (PLASMA MEMBRANE INTRINSIC PROTEIN 2;8); water channel |
| PGWD1-0589 | Canada | PIP2;8 (PLASMA MEMBRANE INTRINSIC PROTEIN 2;8); water channel |
| PGWD1-0613 | Canada | pollen Ole e 1 allergen and extensin family protein |
| PGWD1-0634 | Canada | GMD1 (GDP-D-MANNOSE 4,6-DEHYDRATASE 1); GDP-mannose 4,6-dehydratase/ binding / catalytic/ coenzyme binding |
| PGWD1-0640 | Canada | RSR4 (REDUCED SUGAR RESPONSE 4); protein heterodimerization/ protein homodimerization |
| PGWD1-0667 | Canada | GATL4 (Galacturonosyltransferase-like 4); polygalacturonate 4-alpha-galacturonosyltransferase/ transferase, transferring glycosyl groups / transferase, transferring hexosyl groups |
| PGWD1-0737 | Canada | pectate lyase family protein |
| PGWD1-0787 | Canada | PGSIP1 (PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN 1); transferase, transferring glycosyl groups |
| PGWD1-0788 | Canada | PIP2;8 (PLASMA MEMBRANE INTRINSIC PROTEIN 2;8); water channel |
| PGWD1-0794 | Canada | calcium-binding EF hand family protein |
| PGWD1-0802 | Canada | fringe-related protein |
| PGWD1-0805 | Canada | cupin family protein |
| PGWD1-0807 | Canada | PSAT; O-phospho-L-serine:2-oxoglutarate aminotransferase |
| PGWD1-0808 | Canada | myb family transcription factor |
| PGWD1-0813 | Canada | MYB20 (myb domain protein 20); DNA binding / transcription factor |
| PGWD1-0853 | Canada | TIP4;1 (tonoplast intrinsic protein 4;1); water channel |
| PGWD1-0875 | Canada | GAMMA-TIP (GAMMA TONOPLAST INTRINSIC PROTEIN); water channel |
| PGWD1-0909 | Canada | CUC2 (CUP-SHAPED COTYLEDON 2); transcription factor |
| PGWD1-0932 | Canada | UDP-glucose 6-dehydrogenase, putative |

| | | |
|---|---|---|
| PGWD1-0963 | Canada | MYB36 (myb domain protein 36); DNA binding / transcription factor |
| PGWD1-0965 | Canada | ATHB-3 (ARABIDOPSIS THALIANA HOMEOBOX 3); DNA binding / sequence-specific DNA binding / transcription factor |
| PGWD1-0972 | Canada | cinnamoyl-CoA reductase family |
| PGWD1-1016 | Canada | MYB4; DNA binding / transcription factor |
| PGWD1-1027 | Canada | RAP2.12; DNA binding / transcription factor |
| PGWD1-1034 | Canada | IAA9 (INDOLE-3-ACETIC ACID INDUCIBLE 9); transcription factor |
| PGWD1-1048 | Canada | APE2 (ACCLIMATION OF PHOTOSYNTHESIS TO ENVIRONMENT 2); antiporter/ triose-phosphate transmembrane transporter |
| PGWD1-1070 | Canada | BETA-TIP (BETA-TONOPLAST INTRINSIC PROTEIN); water channel |
| PGWD1-1080 | Canada | TIP4;1 (tonoplast intrinsic protein 4;1); water channel |
| PGWD1-1094 | Canada | LACS9 (LONG CHAIN ACYL-COA SYNTHETASE 9); long-chain-fatty-acid-CoA ligase |
| PGWD1-1121 | Canada | NA |
| PGWD1-1127 | Canada | anac075 (Arabidopsis NAC domain containing protein 75); transcription factor |
| PGWD1-1154 | Canada | PRLI-interacting factor-related |
| PGWD1-1197 | Canada | pectate lyase family protein |
| PGWD1-1206 | Canada | DNAJ heat shock N-terminal domain-containing protein / cell division protein-related |
| PGWD1-1208 | Canada | glycosyl hydrolase family 18 protein |
| PGWD1-1212 | Canada | WRKY7; calmodulin binding / transcription factor |
| PGWD1-1219 | Canada | reversibly glycosylated polypeptide, putative |
| PGWD1-1223 | Canada | pectate lyase family protein |
| PGWD1-1282 | Canada | MYB83 (myb domain protein 83); DNA binding / transcription factor |
| PGWD1-1284 | Canada | G6PD2 (GLUCOSE-6-PHOSPHATE DEHYDROGENASE 2); glucose-6-phosphate dehydrogenase |
| PGWD1-1295 | Canada | unknown protein |
| PGWD1-1304 | Canada | LDOX (LEUCOANTHOCYANIDIN DIOXYGENASE); leucocyanidin oxygenase |
| PGWD1-1319 | Canada | pectate lyase family protein |
| PGWD1-1346 | Canada | MTHFR2 (METHYLENETETRAHYDROFOLATE REDUCTASE 2); methylenetetrahydrofolate reductase (NADPH) |
| PGWD1-1421 | Canada | oxidoreductase, zinc-binding dehydrogenase family protein |
| PGWD1-1422 | Canada | GAMMA-TIP (GAMMA TONOPLAST INTRINSIC PROTEIN); water channel |
| PGWD1-1437 | Canada | TIP4;1 (tonoplast intrinsic protein 4;1); water channel |
| PGWD1-1492 | Canada | glycosyl hydrolase family 18 protein |
| PGWD1-1501 | Canada | polygalacturonase |
| PGWD1-1510 | Canada | pectinesterase family protein |
| PGWD1-1518 | Canada | HCT (HYDROXYCINNAMOYL-COA SHIKIMATE/QUINATE HYDROXYCINNAMOYL TRANSFERASE); quinate O-hydroxycinnamoyltransferase/ shikimate O-hydroxycinnamoyltransferase/ transferase |

## Uppsala dataset (Sweden)

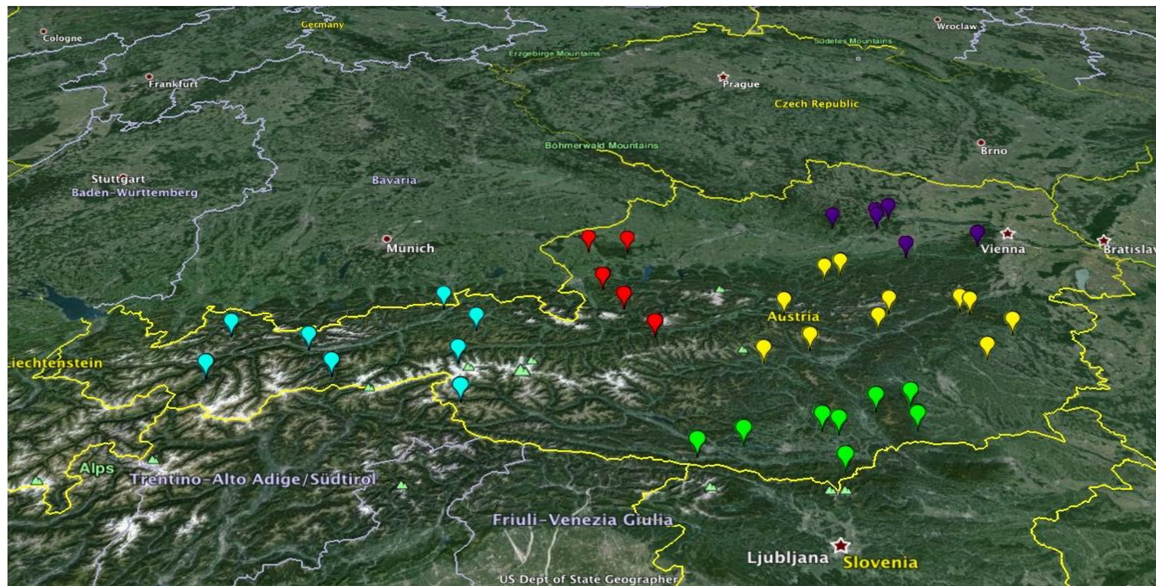
| SNP name | Source | Annotation |
|---|---|---|
| PabiesGI_F6_8_261 | Uppsala | circadian clock |
| PabiesGI_F8_56_324 | Uppsala | circadian clock |
| PaCCA1-Like_1842 | Uppsala | circadian clock |
| PabiesFT4pr_1472 | Uppsala | photoperiodic pathway |
| PabiesFT4pr_1824 | Uppsala | photoperiodic pathway |
| PabiesFT4pr_1951 | Uppsala | photoperiodic pathway |
| PabiesFT4pr_2046 | Uppsala | photoperiodic pathway |
| PabiesFT4pr_2173 | Uppsala | photoperiodic pathway |
| PaFTL2pr_2454 | Uppsala | photoperiodic pathway |
| PaFTL2pr_2509 | Uppsala | photoperiodic pathway |
| PaFTL2pr_2694 | Uppsala | photoperiodic pathway |
| PaCCA1-Like_3194 | Uppsala | Picea abies CCA1-Like gene partial sequence |
| PaCCA1-Like_3893 | Uppsala | Picea abies CCA1-Like gene partial sequence |
| PabiesCol1_711 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesCol1_660 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesPrMYB2_932 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesCol1_846 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesCol1_1495 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesPrMYB2_753 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesMYB2_1302 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesMYB2_2004 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesMYB2_3403 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesMYB2_4049 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesMYB2_4128 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesMYB2_4183 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesPrMYB2_595 | Uppsala | picea abies col1 gene for constans-like 1 |
| PabiesCol2_989 | Uppsala | picea abies constans-like protein gene |
| PaFTL2pr_2790 | Uppsala | Picea abies FTL2 promoter gene |
| PaFTL2pr_1560 | Uppsala | Picea abies FTL2 promoter gene |
| PaFTL2pr_1757 | Uppsala | Picea abies FTL2 promoter gene |
| PabiesKN2b_2317 | Uppsala | Picea abies genotype 5433 homeobox transcription factor KN2 (KN2) gene |
| PabiesHB3_5122 | Uppsala | Picea abies genotype 5433 homeobox transcription factor KN2 (KN2) gene |
| PabiesHB3_385 | Uppsala | Picea abies genotype 5434 homeodomain-leucine zipper trancription factor HB-3 (HB-3) gene |
| PabiesHB3_2495 | Uppsala | Picea abies genotype 5434 homeodomain-leucine zipper trancription factor HB-3 (HB-3) gene |
| PabiesHB3_2316 | Uppsala | Picea abies genotype 5434 homeodomain-leucine zipper trancription factor HB-3 (HB-3) gene |
| PabiesHB3_5700 | Uppsala | Picea abies genotype 5434 homeodomain-leucine zipper trancription factor HB-3 (HB-3) gene |
| PaKN1b_08525f | Uppsala | Picea abies genotype 5460 homeobox transcription factor KN1 (KN1) gene |
| PabiesKN2a_196 | Uppsala | Picea abies genotype 86453 homeobox transcription factor KN2 (KN2) gene |
| PabiesKN2a_253 | Uppsala | Picea abies genotype 86453 homeobox transcription factor KN2 (KN2) gene |

| | | |
|---|---|---|
| PabiesKN4b_489 | Uppsala | Picea abies genotype 86453 homeobox transcription factor KN4 (KN4) gene |
| PaKN4b_01687n | Uppsala | Picea abies genotype 86453 homeobox transcription factor KN4 (KN4) gene |
| PaMFTL1_2136 | Uppsala | Picea abies MFT1-like protein (MFT1) gene |
| PabiesFT1_1718 | Uppsala | Picea abies MFT1-like protein (MFT1) gene |
| PaMFTL1_1251 | Uppsala | Picea abies MFT1-like protein (MFT1) gene |
| PaMFTL1_2215 | Uppsala | Picea abies MFT1-like protein (MFT1) gene |
| PaMFTL1_1613 | Uppsala | Picea abies MFT1-like protein (MFT1) gene |
| PabiesFT1_1050 | Uppsala | Picea abies MFT1-like protein (MFT1) gene |
| PabiesFT1_2091 | Uppsala | Picea abies MFT1-like protein (MFT1) gene |
| PabiesFT1_911 | Uppsala | Picea abies MFT1-like protein (MFT1) gene |
| PabiesCry_454 | Uppsala | Picea abies partial cry gene for putative cryptochrome |
| PaPHYN_R0111 | Uppsala | Picea abies partial phynrI gene for phytochrome N |
| PaPHYN_RIII418 | Uppsala | Picea abies partial phynrI gene for phytochrome N |
| PaPHYN_RIII272 | Uppsala | Picea abies partial phynrI gene for phytochrome N |
| PaPHYN_RII465 | Uppsala | Picea abies partial phynrI gene for phytochrome N |
| PaPHYN_RI330 | Uppsala | Picea abies partial phynrI gene for phytochrome N |
| PaPHYN_RIII185 | Uppsala | Picea abies partial phynrI gene for phytochrome N |
| PaPHYN_RIII88 | Uppsala | Picea abies partial phynrI gene for phytochrome N |
| PaPHYN_R0319 | Uppsala | Picea abies partial phynrI gene for phytochrome N |
| PaPHYO_RIII336 | Uppsala | Picea abies partial phyo gene for phytochrome O |
| PaPHYO_RIV211 | Uppsala | Picea abies partial phyo gene for phytochrome O |
| PaPHYO_RIII510 | Uppsala | Picea abies partial phyo gene for phytochrome O |
| PaPHYO_RII229 | Uppsala | Picea abies partial phyo gene for phytochrome O |
| PaPHYO_RIV39 | Uppsala | Picea abies partial phyo gene for phytochrome O |
| PaPHYO_RII283 | Uppsala | Picea abies partial phyo gene for phytochrome O |
| PaPHYO_RIV402 | Uppsala | Picea abies partial phyo gene for phytochrome O |
| PaPHYP_RII122 | Uppsala | Picea abies partial phyP gene for putative phytochrome P, |
| PaPHYP_RII177 | Uppsala | Picea abies partial phyP gene for putative phytochrome P, |
| PaPHYP_RI96 | Uppsala | Picea abies partial phyP gene for putative phytochrome P, |
| PaPHYP_RIII76 | Uppsala | Picea abies partial phyP gene for putative phytochrome P, |
| PaPHYP_RIII345 | Uppsala | Picea abies partial phyP gene for putative phytochrome P, |
| PaPHYP_RIII274 | Uppsala | Picea abies partial phyP gene for putative phytochrome P, |
| PabiesGI_F2_9_1470 | Uppsala | Picea abies Putative Gigantea gene,partial sequence |

| | | |
|---|---|---|
| PabiesGI_F2_9_420 | Uppsala | Picea abies Putative Gigantea gene,partial sequence |
| PabiesPRR1_240 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR1_2953 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR1_3828 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR1_1039 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR1_3301 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR1_1168 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR1_3722 | Uppsala | Picea abies Putative PRR1 gene |
| PaPRR1_3_GQ0178.B 7-E07.1-180 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR1_2381 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR1_2741 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR1_2920 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR1_2990 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR1_3883 | Uppsala | Picea abies Putative PRR1 gene |
| PabiesPRR3_F2_481 | Uppsala | Picea abies Putative PRR3 gene, partial sequence |
| PabiesPRR3_F1_2570 | Uppsala | Picea abies Putative PRR3 gene, partial sequence |
| PabiesPRR3_F1_2978 | Uppsala | Picea abies Putative PRR3 gene, partial sequence |
| PabiesPRR3_F2_331 | Uppsala | Picea abies Putative PRR3 gene, partial sequence |
| PabiesPRR7_F3_104 | Uppsala | Picea abies Putative PRR7 gene,partial sequence |
| PabiesPRR7_F1_1505 | Uppsala | Picea abies Putative PRR7 gene,partial sequence |
| PabiesPRR7_F2_534 | Uppsala | Picea abies Putative PRR7 gene,partial sequence |
| PabiesPRR7_F1_2518 | Uppsala | Picea abies Putative PRR7 gene,partial sequence |
| PabiesPRR7_F1_771 | Uppsala | Picea abies Putative PRR7 gene,partial sequence |
| PabiesPRR7_F2_417 | Uppsala | Picea abies Putative PRR7 gene,partial sequence |
| PabiesZTL_793 | Uppsala | Picea likiangensis isolate PLJT-TBS-2 putative ZTL (ztl) gene, partial cds |
| PabiesZTL_514 | Uppsala | Picea likiangensis isolate PLJT-TBS-2 putative ZTL (ztl) gene, partial cds |
| PabiesZTL_958 | Uppsala | Picea likiangensis isolate PLJT-TBS-2 putative ZTL (ztl) gene, partial cds |
| PabiesZTL_367 | Uppsala | Picea likiangensis isolate PLJT-TBS-2 putative ZTL (ztl) gene, partial cds |
| PaMFTL1_3441 | Uppsala | shoot apical development |
| PaMFTL1_802 | Uppsala | shoot apical development |

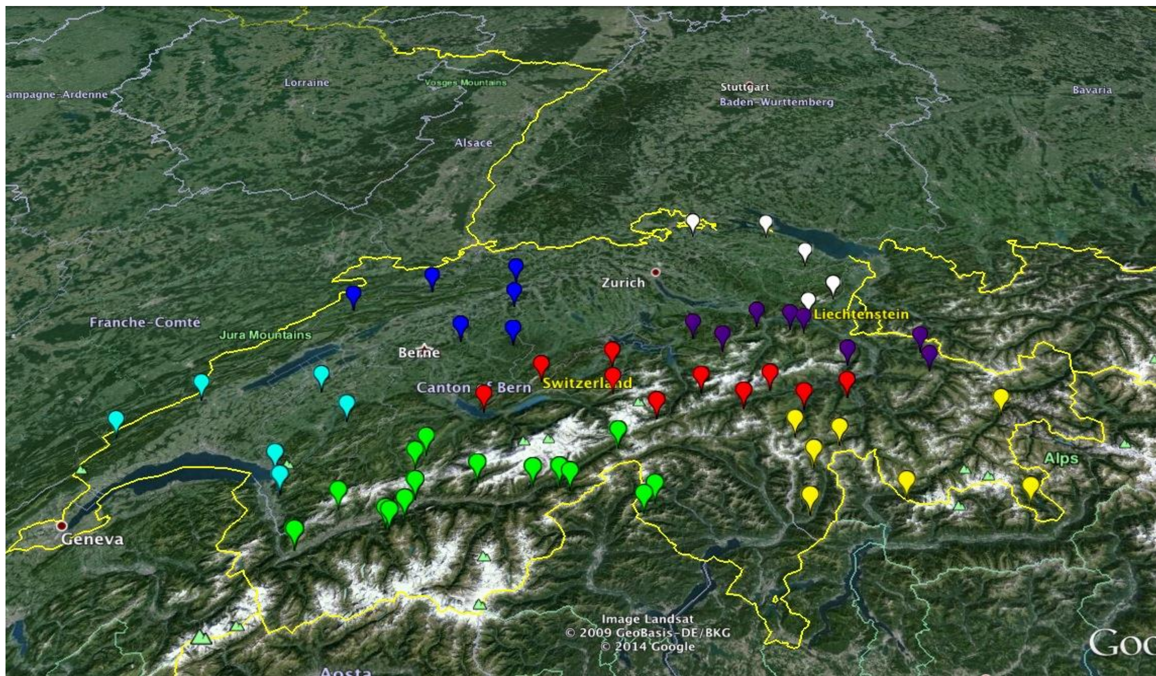
Supplement 3. Assignment of 392 individuals based on the geo-reference data when individuals from Austria were assigned in the five populations (A), from Italy-Slovenia and France in 8 populations (B) and from Switzerland, individuals are assigned in the 7 adjusted populations (C).

(A) Assignment of individuals in 5 adjusted groups




(B) Assignment of individuals in 8 adjusted groups

(C) Assignment of individuals in 7 adjusted groups



Supplement 4. List of SNPs failed during 384 Illumina GoldenGate genotyping assay

| **SNPs name** | **Source** | **Species** | **Final score** | **Design-ability score** |
|---|---|---|---|---|
| PGWD1-1437 | Canada | *P.glauca* | 0.626 | 1 |
| PGWD1-1048 | Canada | *P.glauca* | 0.869 | 1 |
| PGWD1-0257 | Canada | *P.glauca* | 0.956 | 1 |
| PGWD1-0965 | Canada | *P.glauca* | 0.947 | 1 |
| PGLM2-0195 | Canada | *P.glauca* | 0.995 | 1 |
| PGWD1-1510 | Canada | *P.glauca* | 0.993 | 1 |
| PGLM2-0360 | Canada | *P.glauca* | 0.978 | 1 |
| PGLM2-1273 | Canada | *P.glauca* | 0.706 | 1 |
| PGLM2-0349 | Canada | *P.glauca* | 0.977 | 1 |
| PGLM2-0924 | Canada | *P.glauca* | 0.956 | 1 |
| PGWD1-0963 | Canada | *P.glauca* | 0.958 | 1 |
| PGLM2-0733 | Canada | *P.glauca* | 0.955 | 1 |
| PGWD1-0613 | Canada | *P.glauca* | 0.937 | 1 |
| PGWD1-0519 | Canada | *P.glauca* | 0.971 | 1 |
| PGLM2-0465 | Canada | *P.glauca* | 0.927 | 1 |
| PGWD1-1027 | Canada | *P.glauca* | 0.911 | 1 |
| PGLM2-1223 | Canada | *P.glauca* | 0.931 | 1 |
| PGLM2-1171 | Canada | *P.glauca* | 0.918 | 1 |
| PGWD1-1422 | Canada | *P.glauca* | 0.609 | 1 |
| PGLM2-0314 | Canada | *P.glauca* | 0.96 | 1 |
| PGLM2-0383 | Canada | *P.glauca* | 0.932 | 1 |
| PGLM2-0810 | Canada | *P.galuca* | 0.951 | 1 |
| PGWD1-0972 | Canada | *P.glauca* | 0.947 | 1 |
| PGLM2-0227 | Canada | *P.glauca* | 0.993 | 1 |
| PGWD1-0557 | Canada | *P.glauca* | 0.818 | 1 |
| PGLM2-0887 | Canada | *P.glauca* | 0.82 | 1 |
| PGWD1-0578 | Canada | *P.glauca* | 0.968 | 1 |
| PGLM2-1147 | Canada | *P.glauca* | 0.913 | 1 |
| PGWD1-0053 | Canada | *P.glauca* | 0.99 | 1 |
| PGLM2-0780 | Canada | *P.glauca* | 0.946 | 1 |
| PGLM2-0450 | Canada | *P.glauca* | 0.986 | 1 |
| PGWD1-0556 | Canada | *P.glauca* | 0.917 | 1 |
| PGWD1-1208 | Canada | *P.glauca* | 0.944 | 1 |
| PGLM2-0024 | Canada | *P.glauca* | 0.924 | 1 |
| PGWD1-0468 | Canada | *P.glauca* | 0.969 | 1 |
| PGLM2-0271 | Canada | *P.glauca* | 0.911 | 1 |
| PGLM2-0368 | Canada | *P.glauca* | 0.989 | 1 |
| PGWD1-0808 | Canada | *P.glauca* | 0.918 | 1 |
| PGLM2-0762 | Canada | *P.glauca* | 0.964 | 1 |
| PGWD1-0640 | Canada | *P.glauca* | 0.957 | 1 |
| PGWD1-1121 | Canada | *P.glauca* | 0.931 | 1 |
| PGWD1-1197 | Canada | *P.glauca* | 0.912 | 1 |
| PGLM2-0796 | Canada | *P.glauca* | 0.975 | 1 |
| PGWD1-0485 | Canada | *P.glauca* | 0.945 | 1 |
| PGLM2-0013 | Canada | *P.glauca* | 0.918 | 1 |
| PGWD1-0533 | Canada | *P.glauca* | 0.977 | 1 |

| PGLM2-1037 | Canada | *P.glauca* | 0.961 | 1 |
|---|---|---|---|---|
| PGLM2-0674 | Canada | *P.glauca* | 0.956 | 1 |
| PGLM2-0049 | Canada | *P.glauca* | 0.989 | 1 |
| PGWD1-0051 | Canada | *P.glauca* | 0.98 | 1 |
| PGWD1-0551 | Canada | *P.glauca* | 0.98 | 1 |
| PGLM2-1492 | Canada | *P.glauca* | 0.935 | 1 |
| PGWD1-1094 | Canada | *P.glauca* | 0.93 | 1 |
| PGWD1-0413 | Canada | *P.glauca* | 0.965 | 1 |
| PaPHYN_R0319 | Canada | *P.glauca* | 0.879 | 1 |
| PGWD1-0117 | Canada | *P.glauca* | 0.932 | 1 |
| PGLM2-0296 | Canada | *P.glauca* | 0.921 | 1 |
| PGWD1-0875 | Canada | *P.glauca* | 0.781 | 1 |
| PGWD1-0932 | Canada | *P.glauca* | 0.945 | 1 |
| PGWD1-1080 | Canada | *P.glauca* | 0.824 | 1 |
| PGWD1-0242 | Canada | *P.glauca* | 0.985 | 1 |
| PGWD1-0142 | Canada | *P.glauca* | 0.917 | 1 |
| PGWD1-1206 | Canada | *P.glauca* | 0.888 | 1 |
| 0_8531_01-Paab_157 | CRSP reseq | *P.abies* | 0.979 | 1 |
| 2_4196_01-Paab_201 | CRSP reseq | *P.abies* | 0.629 | 1 |
| 0_4541_02-266 | CRSP reseq | *P.abies* | 0.97 | 1 |
| 2_10438_01-351 | CRSP reseq | *P.abies* | 0.818 | 1 |
| CL4511Contig1_02-Paab_223 | CRSP reseq | *P.abies* | 0.956 | 1 |
| CL1308Contig1_03-Paab_181 | CRSP reseq | *P.abies* | 0.993 | 1 |
| CL1343Contig1_05-Paab_165 | CRSP reseq | *P.abies* | 0.972 | 1 |
| 0_10631_01-Paab_193 | CRSP reseq | *P.abies* | 0.64 | 1 |
| 0_7471_01-Paab_399 | CRSP reseq | *P.abies* | 0.88 | 1 |
| 0_8111_01-39 | CRSP reseq | *P.abies* | 0.919 | 1 |
| PaFTL2pr_2790 | Uppsala | *P.abies* | 0.908 | 1 |
| PaPHYP_RI96 | Uppsala | *P.abies* | 0.954 | 1 |
| PGWD1-0007 | Uppsala | *P.abies* | 0.909 | 1 |
| PaPRR1_3_GQ0178.B7-E07.1-180 | Uppsala | *P.abies* | 0.62 | 1 |
| PaCCA1-Like_3194 | Uppsala | *P.abies* | 0.767 | 1 |
| PabiesPrMYB2_753 | Uppsala | *P.abies* | 0.981 | 1 |
| PabiesPRR7_F2_534 | Uppsala | *P.abies* | 0.886 | 1 |
| PaPHYP_RII122 | Uppsala | *P.abies* | 0.84 | 1 |
| PabiesPRR1_3301 | Uppsala | *P.abies* | 0.775 | 1 |
| PaPHYP_RIII76 | Uppsala | *P.abies* | 0.64 | 1 |
| PaPHYN_RIII418 | Uppsala | *P.abies* | 0.859 | 1 |
| PabiesCol1_583 | Uppsala | *P.abies* | 0.607 | 1 |
| PabiesMYB2_2004 | Uppsala | *P.abies* | 0.988 | 1 |
| PabiesPRR1_2741 | Uppsala | *P.abies* | 0.843 | 1 |
| PabiesPRR1_2920 | Uppsala | *P.abies* | 0.947 | 1 |
| PabiesCol1_846 | Uppsala | *P.abies* | 0.937 | 1 |
| PabiesMYB2_1302 | Uppsala | *P.abies* | 0.77 | 1 |
| PabiesHB3_5122 | Uppsala | *P.abies* | 0.962 | 1 |
| PaPHYP_RIII345 | Uppsala | *P.abies* | 0.998 | 1 |
| PabiesFT4pr_1951 | Uppsala | *P.abies* | 0.786 | 1 |
| PaPHYP_RIII274 | Uppsala | *P.abies* | 0.923 | 1 |

| | | | | |
|---|---|---|---|---|
| PabiesKN2a_196 | Uppsala | *P.abies* | 0.672 | 1 |
| PabiesFT4pr_2173 | Uppsala | *P.abies* | 0.746 | 1 |
| PabiesPrMYB2_595 | Uppsala | *P.abies* | 0.867 | 1 |
| PabiesFT4pr_1824 | Uppsala | *P.abies* | 0.902 | 1 |
| PaPHYO_RIV402 | Uppsala | *P.abies* | 0.813 | 1 |
| PabiesPRR1_3722 | Uppsala | *P.abies* | 0.62 | 1 |

Supplement 5. Basic diversity statistics indices on genotyping matrix on polymorphic SNPs. Listed observed and expected heterozygosity for each SNP when maf ≥ 0.5 and Fis across all loci.

| SNP | Ho |
|---|---|
| CL1148Contig1_08.Paab_225 | 0.17829457 |
| PGWD1.0396 | 0.48818898 |
| PGWD1.1319 | 0.94805195 |
| PGWD1.0282 | 0.13350785 |
| CL1530Contig1_04.Paab_64 | 0.27109974 |
| X0_9457_01.Paab_115 | 0.49739583 |
| PGWD1.1501 | 0.4 |
| CL1694Contig1_01.Paab_235 | 0.43121693 |
| PGLM2.0886 | 0.40220386 |
| PGWD1.0667 | 0.11658031 |
| X0_8531_01.Paab_363 | 0.30343008 |
| X0_9749_01.Paab_337 | 0.24675325 |
| CL1148Contig1_08.Paab_134 | 0.41239892 |
| PGWD1.0634 | 0.42506812 |
| CL1694Contig1_04.Paab_90 | 0.43421053 |
| X0_366_02.Paab_380 | 0.18324607 |
| PGWD1.1034 | 0.16397849 |
| PGWD1.0337 | 0.73385013 |
| X0_9457_01.Paab_46 | 0.5078125 |
| CL4284Contig1_01.Paab_180 | 0.4488189 |
| PaCCA1.Like_3893 | 0.1865285 |
| PGLM2.0645 | 0.32992327 |
| PGWD1.0807 | 0.41578947 |
| CL304Contig1_01.Paab_118 | 0.41732283 |
| PGWD1.0391 | 0.48266667 |
| PGLM2.0127 | 0.47164948 |
| PGWD1.1223 | 0.45646438 |
| X2_9328_01.Paab_425 | 0.41494845 |

| SNP | He |
|---|---|
| CL1148Contig1_08.Paab_225 | 0.20803377 |
| PGWD1.0396 | 0.48390662 |
| PGWD1.1319 | 0.5002063 |
| PGWD1.0282 | 0.1249227 |
| CL1530Contig1_04.Paab_64 | 0.23867795 |
| X0_9457_01.Paab_115 | 0.50007819 |
| PGWD1.1501 | 0.42731565 |
| CL1694Contig1_01.Paab_235 | 0.40111644 |
| PGLM2.0886 | 0.40996606 |
| PGWD1.0667 | 0.11464572 |
| X0_8531_01.Paab_363 | 0.43676969 |
| X0_9749_01.Paab_337 | 0.21687297 |
| CL1148Contig1_08.Paab_134 | 0.44352371 |
| PGWD1.0634 | 0.48465627 |
| CL1694Contig1_04.Paab_90 | 0.43424177 |
| X0_366_02.Paab_380 | 0.18382323 |
| PGWD1.1034 | 0.30802756 |
| PGWD1.0337 | 0.4657857 |
| X0_9457_01.Paab_46 | 0.50103011 |
| CL4284Contig1_01.Paab_180 | 0.4316791 |
| PaCCA1.Like_3893 | 0.19036404 |
| PGLM2.0645 | 0.33675979 |
| PGWD1.0807 | 0.40333287 |
| CL304Contig1_01.Paab_118 | 0.41290579 |
| PGWD1.0391 | 0.36991444 |
| PGLM2.0127 | 0.44926277 |
| PGWD1.1223 | 0.4755099 |
| X2_9328_01.Paab_425 | 0.42129186 |

| SNP | Fis |
|---|---|
| PGWD1.1319 | -0.89532 |
| PGWD1.0337 | -0.57551 |
| PGLM2.0353 | -0.40109 |
| X0_9457_01.Paab_421 | -0.37184 |
| PGWD1.0391 | -0.30481 |
| PGWD1.1421 | -0.27803 |
| PGLM2.0169 | -0.13983 |
| X0_9749_01.Paab_337 | -0.13778 |
| CL1530Contig1_04.Paab_64 | -0.13584 |
| PGLM2.0158 | -0.11699 |
| PabiesCol1_1495 | -0.09658 |
| PabiesZTL_514 | -0.09283 |
| PGLM2.1166 | -0.08264 |
| PGLM2.1170 | -0.07720 |
| PabiesZTL_793 | -0.07579 |
| CL1694Contig1_01.Paab_235 | -0.07504 |
| PabiesZTL_958 | -0.07391 |
| X0_9383_01.Paab_438 | -0.07324 |
| X2_5636_01.Paab_209 | -0.06955 |
| PGWD1.0282 | -0.06872 |
| PGLM2.0140 | -0.06777 |
| PGLM2.1030 | -0.06673 |
| CL1694Contig1_02.365 | -0.06577 |
| PGLM2.0489 | -0.06526 |
| PabiesKN2b_2317 | -0.06505 |
| PGWD1.0421 | -0.06296 |
| PGLM2.0642 | -0.06285 |
| X0_11090_01.Paab_251 | -0.05961 |

| | |
|---|---|
| X0_1439_01.Paab_226 | 0.22135417 |
| PGWD1.0589 | 0.10789474 |
| PabiesPRR1_2953 | 0.36458333 |
| PGLM2.0901 | 0.1761658 |
| PGLM2.1269 | 0.09511568 |
| PabiesFT1_1050 | 0.34574468 |
| PGLM2.0169 | 0.50793651 |
| X0_177_01.Paab_165 | 0.45736434 |
| PGWD1.0029 | 0.18298969 |
| PGLM2.1031 | 0.17268041 |
| PGWD1.0184 | 0.15721649 |
| PGLM2.0433 | 0.48812665 |
| PGLM2.0098 | 0.26030928 |
| PaPHYO_RIII510 | 0.27055703 |
| PGWD1.0853 | 0.35752688 |
| PGLM2.0353 | 0.57622739 |
| PGLM2.0160 | 0.44125326 |
| UMN_1604_01.Paab_348 | 0.25510204 |
| CL3795Contig1_01.Paab_45 | 0.49041096 |
| PGLM2.0584 | 0.27466667 |
| PGWD1.0132 | 0.36814621 |
| PGWD1.0787 | 0.13554987 |
| PGWD1.1421 | 0.46498599 |
| PGLM2.0437 | 0.25064599 |
| PGLM2.0140 | 0.13178295 |
| PabiesHB3_2495 | 0.18586387 |
| X0_489_01.Paab_316 | 0.36528497 |
| PabiesZTL_367 | 0.13265306 |
| PGWD1.0041 | 0.18974359 |
| PGLM2.0285 | 0.47135417 |
| PGLM2.0391 | 0.47860963 |
| X2_3947_01.Paab_298 | 0.1056701 |
| PGWD1.1346 | 0.2377261 |

| | |
|---|---|
| X0_1439_01.Paab_226 | 0.21735952 |
| PGWD1.0589 | 0.11634148 |
| PabiesPRR1_2953 | 0.35422106 |
| PGLM2.0901 | 0.18625934 |
| PGLM2.1269 | 0.11381616 |
| PabiesFT1_1050 | 0.35948936 |
| PGLM2.0169 | 0.44562334 |
| X0_177_01.Paab_165 | 0.49224471 |
| PGWD1.0029 | 0.17922361 |
| PGLM2.1031 | 0.17506793 |
| PGWD1.0184 | 0.16667666 |
| PGLM2.0433 | 0.46357373 |
| PGLM2.0098 | 0.2744579 |
| PaPHYO_RIII510 | 0.26104746 |
| PGWD1.0853 | 0.4500442 |
| PGLM2.0353 | 0.4112711 |
| PGLM2.0160 | 0.42441527 |
| UMN_1604_01.Paab_348 | 0.25636776 |
| CL3795Contig1_01.Paab_45 | 0.49955216 |
| PGLM2.0584 | 0.3160107 |
| PGWD1.0132 | 0.36732943 |
| PGWD1.0787 | 0.13110696 |
| PGWD1.1421 | 0.36383093 |
| PGLM2.0437 | 0.25366175 |
| PGLM2.0140 | 0.12341848 |
| PabiesHB3_2495 | 0.19005854 |
| X0_489_01.Paab_316 | 0.37597403 |
| PabiesZTL_367 | 0.14602798 |
| PGWD1.0041 | 0.19268341 |
| PGLM2.0285 | 0.50073093 |
| PGLM2.0391 | 0.49789609 |
| X2_3947_01.Paab_298 | 0.11409135 |
| PGWD1.1346 | 0.22942523 |

| | |
|---|---|
| PaPHYO_RIII336 | -0.05786 |
| PGLM2.0154 | -0.05735 |
| CL4257Contig1_01.Paab_391 | -0.05331 |
| PGLM2.0433 | -0.05296 |
| PabiesCol1_660 | -0.05292 |
| PGLM2.0127 | -0.04983 |
| PaFTL2pr_1560 | -0.04846 |
| X0_13680_01.Paab_216 | -0.04592 |
| PGLM2.0404 | -0.04484 |
| PabiesKN4b_489 | -0.04002 |
| CL4284Contig1_01.Paab_180 | -0.03970 |
| PGLM2.0160 | -0.03967 |
| X2_7725_01.Paab_466 | -0.03948 |
| PaPHYN_RI330 | -0.03764 |
| PaPHYO_RIII510 | -0.03643 |
| PaPHYP_RII177 | -0.03619 |
| PGWD1.1346 | -0.03618 |
| X2_9280_01.Paab_338 | -0.03581 |
| PGWD1.1127 | -0.03469 |
| PabiesPRR1_2990 | -0.03468 |
| X0_13978_01.Paab_102 | -0.03460 |
| CL1692Contig1_05.Paab_178 | -0.03423 |
| PGWD1.0787 | -0.03389 |
| PGLM2.0395 | -0.03381 |
| PGWD1.0807 | -0.03088 |
| PGLM2.1182 | -0.02953 |
| PabiesPRR1_2953 | -0.02925 |
| X0_17215_01.Paab_108 | -0.02850 |
| PabiesFT4pr_2046 | -0.02799 |
| PabiesPRR7_F1_771 | -0.02672 |
| PGWD1.1304 | -0.02644 |
| X2_9845_01.Paab_282 | -0.02460 |
| PGWD1.0029 | -0.02101 |

| | |
|---|---|
| CL3771Contig1_04.Paab_68 | 0.34173669 |
| PGLM2.0923 | 0.30287206 |
| CL3582Contig1_03.Paab_63 | 0.47382199 |
| PGWD1.1492 | 0.43832021 |
| PabiesHB3_385 | 0.15979381 |
| CL1694Contig1_02.365 | 0.42708333 |
| PGWD1.1154 | 0.11734694 |
| X0_12021_01.161 | 0.14578005 |
| PGWD1.0042 | 0.27393617 |
| PabiesPRR7_F1_771 | 0.18911917 |
| PGLM2.1170 | 0.45406824 |
| PGLM2.0565 | 0.10209424 |
| CL304Contig1_01.202 | 0.39037433 |
| PaMFTL1_1613 | 0.41145833 |
| PabiesZTL_793 | 0.43569554 |
| CL1692Contig1_05.Paab_178 | 0.51168831 |
| PabiesZTL_514 | 0.43264249 |
| PGLM2.0081 | 0.10309278 |
| PGLM2.1182 | 0.43617021 |
| PGWD1.0024 | 0.10103627 |
| X2_5636_01.Paab_399 | 0.33506494 |
| CL4257Contig1_01.Paab_391 | 0.27083333 |
| X0_9457_01.Paab_421 | 0.54593176 |
| UMN_3055_01.Paab_224 | 0.16795866 |
| PaKN4b_01687n | 0.19845361 |
| PabiesFT1_1718 | 0.41836735 |
| PGWD1.0400 | 0.2997416 |
| X0_7921_01.Paab_212 | 0.14138817 |
| X0_13680_01.Paab_216 | 0.51570681 |
| PGLM2.1030 | 0.37665782 |
| PabiesPRR3_F2_481 | 0.47311828 |
| PGWD1.0421 | 0.26493506 |
| PGLM2.1166 | 0.28608924 |

| | |
|---|---|
| CL3771Contig1_04.Paab_68 | 0.36087244 |
| PGLM2.0923 | 0.3433352 |
| CL3582Contig1_03.Paab_63 | 0.49015061 |
| PGWD1.1492 | 0.46814822 |
| PabiesHB3_385 | 0.16878446 |
| CL1694Contig1_02.365 | 0.40072889 |
| PGWD1.1154 | 0.1197218 |
| X0_12021_01.161 | 0.14420946 |
| PGWD1.0042 | 0.28147163 |
| PabiesPRR7_F1_771 | 0.18419689 |
| PGLM2.1170 | 0.42152576 |
| PGLM2.0565 | 0.11115348 |
| CL304Contig1_01.202 | 0.42283265 |
| PaMFTL1_1613 | 0.45070442 |
| PabiesZTL_793 | 0.40500069 |
| CL1692Contig1_05.Paab_178 | 0.49475446 |
| PabiesZTL_514 | 0.39589193 |
| PGLM2.0081 | 0.10724846 |
| PGLM2.1182 | 0.42365957 |
| PGWD1.0024 | 0.10083776 |
| X2_5636_01.Paab_399 | 0.34057427 |
| CL4257Contig1_01.Paab_391 | 0.25712576 |
| X0_9457_01.Paab_421 | 0.39795552 |
| UMN_3055_01.Paab_224 | 0.20803377 |
| PaKN4b_01687n | 0.20359826 |
| PabiesFT1_1718 | 0.43363432 |
| PGWD1.0400 | 0.30038425 |
| X0_7921_01.Paab_212 | 0.14488975 |
| X0_13680_01.Paab_216 | 0.49306386 |
| PGLM2.1030 | 0.35309555 |
| PabiesPRR3_F2_481 | 0.49889865 |
| PGWD1.0421 | 0.24924242 |
| PGLM2.1166 | 0.26425266 |

| | |
|---|---|
| X0_17587_01.Paab_42 | -0.02095 |
| CL3507Contig1_03.Paab_191 | -0.01872 |
| PGWD1.1219 | -0.01842 |
| X0_1439_01.Paab_226 | -0.01838 |
| PGWD1.0667 | -0.01687 |
| PGWD1.0344 | -0.01641 |
| PabiesPRR7_F1_1505 | -0.01631 |
| PaPHYO_RII283 | -0.01416 |
| X0_9457_01.Paab_46 | -0.01354 |
| PGWD1.1070 | -0.01295 |
| X2_9665_01.Paab_175 | -0.01098 |
| X0_12021_01.161 | -0.01089 |
| CL304Contig1_01.Paab_118 | -0.01070 |
| PGWD1.0118 | -0.00936 |
| PGWD1.0396 | -0.00885 |
| PGLM2.0592 | -0.00841 |
| PGWD1.0510 | -0.00471 |
| PGLM2.0571 | -0.00273 |
| X0_2354_01.Paab_194 | -0.00235 |
| PGWD1.0132 | -0.00222 |
| PGWD1.0024 | -0.00197 |
| CL1694Contig1_04.Paab_90 | 0.00007 |
| PGWD1.0147 | 0.00095 |
| PGLM2.0818 | 0.00102 |
| PGLM2.1091 | 0.00108 |
| PGWD1.0400 | 0.00214 |
| X0_366_02.Paab_380 | 0.00314 |
| PGLM2.0610 | 0.00354 |
| UMN_4748_01.Paab_38 | 0.00391 |
| X0_14976_01.Paab_305 | 0.00474 |
| UMN_1604_01.Paab_348 | 0.00494 |
| PabiesHB3_5700 | 0.00512 |
| PabiesHB3_2316 | 0.00523 |

| PGWD1.0158 | 0.46933333 |
|---|---|
| X2_4723_01.Paab_276 | 0.14948454 |
| X2_9466_01.Paab_179 | 0.29473684 |
| X0_14976_01.Paab_305 | 0.49481865 |
| X0_2433_01.Paab_290 | 0.34036939 |
| PGWD1.0147 | 0.09793814 |
| PGLM2.0642 | 0.33766234 |
| PabiesPRR1_3883 | 0.15284974 |
| CL1758Contig1_04.288 | 0.36787565 |
| PabiesCol1_1495 | 0.18108108 |
| X2_6491_01.Paab_360 | 0.46648794 |
| PGWD1.0737 | 0.48958333 |
| PabiesZTL_958 | 0.42408377 |
| PGWD1.0220 | 0.2025974 |
| X2_4892_01.Paab_39 | 0.46194226 |
| PGLM2.0489 | 0.17571059 |
| PabiesPRR1_2990 | 0.5128866 |
| PaMFTL1_2136 | 0.18814433 |
| PGLM2.0784 | 0.44186047 |
| PabiesKN4b_489 | 0.31770833 |
| PGLM2.0021 | 0.44973545 |
| X0_17587_01.Paab_42 | 0.36702128 |
| PGWD1.0362 | 0.40837696 |
| PabiesKN2b_2317 | 0.53367876 |
| PabiesPRR3_F2_331 | 0.48302872 |
| PabiesPRR1_3828 | 0.38845144 |
| PabiesPRR7_F1_1505 | 0.41005291 |
| PGWD1.1127 | 0.43617021 |
| PaPHYN_RI330 | 0.51036269 |
| PaPHYO_RIII336 | 0.24528302 |
| X2_3851_01.Paab_280 | 0.33676093 |
| PaPHYO_RI145 | 0.18670077 |
| PGLM2.0134 | 0.18251928 |

| PGWD1.0158 | 0.48233868 |
|---|---|
| X2_4723_01.Paab_276 | 0.15173553 |
| X2_9466_01.Paab_179 | 0.29800028 |
| X0_14976_01.Paab_305 | 0.49717718 |
| X0_2433_01.Paab_290 | 0.36171839 |
| PGWD1.0147 | 0.09803138 |
| PGLM2.0642 | 0.31769481 |
| PabiesPRR1_3883 | 0.16745508 |
| CL1758Contig1_04.288 | 0.38238342 |
| PabiesCol1_1495 | 0.16513221 |
| X2_6491_01.Paab_360 | 0.491222 |
| PGWD1.0737 | 0.49737544 |
| PabiesZTL_958 | 0.39489632 |
| PGWD1.0220 | 0.21095779 |
| X2_4892_01.Paab_39 | 0.49493024 |
| PGLM2.0489 | 0.16494625 |
| PabiesPRR1_2990 | 0.49569448 |
| PaMFTL1_2136 | 0.19153081 |
| PGLM2.0784 | 0.44645607 |
| PabiesKN4b_489 | 0.30548303 |
| PGLM2.0021 | 0.49962809 |
| X0_17587_01.Paab_42 | 0.35948936 |
| PGWD1.0362 | 0.42808262 |
| PabiesKN2b_2317 | 0.50108337 |
| PabiesPRR3_F2_331 | 0.49899867 |
| PabiesPRR1_3828 | 0.41400746 |
| PabiesPRR7_F1_1505 | 0.40347424 |
| PGWD1.1127 | 0.4215461 |
| PaPHYN_RI330 | 0.49184779 |
| PaPHYO_RIII336 | 0.23186785 |
| X2_3851_01.Paab_280 | 0.34095155 |
| PaPHYO_RI145 | 0.21399764 |
| PGLM2.0134 | 0.21883033 |

| X0_9457_01.Paab_115 | 0.00536 |
|---|---|
| X0_17215_01.Paab_225 | 0.00621 |
| PabiesPRR1_1168 | 0.00786 |
| PGLM2.0784 | 0.01029 |
| X2_9466_01.Paab_179 | 0.01095 |
| X0_7171_01.Paab_359 | 0.01154 |
| PGLM2.0437 | 0.01189 |
| X2_3851_01.Paab_280 | 0.01229 |
| PaMFTL1_2215 | 0.01232 |
| PabiesGI_F8_56_324 | 0.01271 |
| X0_13957_02.Paab_27 | 0.01305 |
| X2_7803_01.Paab_235 | 0.01340 |
| PabiesFT1_911 | 0.01357 |
| PGLM2.1031 | 0.01364 |
| PabiesGI_F2_9_1470 | 0.01463 |
| X2_4723_01.Paab_276 | 0.01483 |
| X2_9328_01.Paab_425 | 0.01506 |
| PGWD1.0041 | 0.01526 |
| PGWD1.0737 | 0.01567 |
| X2_5636_01.Paab_399 | 0.01618 |
| PaMFTL1_2136 | 0.01768 |
| PGWD1.0418 | 0.01798 |
| PGWD1.0794 | 0.01826 |
| CL3795Contig1_01.Paab_45 | 0.01830 |
| PGLM2.0886 | 0.01893 |
| PGWD1.1154 | 0.01984 |
| PaCCA1.Like_3893 | 0.02015 |
| PGLM2.0645 | 0.02030 |
| PabiesPRR3_F1_2570 | 0.02067 |
| PabiesCol1_711 | 0.02131 |
| PabiesCry_454 | 0.02185 |
| PabiesHB3_2495 | 0.02207 |
| X2_9280_01.Paab_123 | 0.02367 |

| | |
|---|---|
| UMN_4748_01.Paab_38 | 0.25848564 |
| PabiesHB3_2316 | 0.45844504 |
| PGWD1.0909 | 0.33695652 |
| X0_11090_01.Paab_251 | 0.22454308 |
| PabiesGI_F8_56_324 | 0.26165803 |
| PGLM2.1528 | 0.34210526 |
| PaPHYO_RIV211 | 0.09793814 |
| PGLM2.1477 | 0.15789474 |
| CL3771Contig1_04.Paab_419 | 0.12532637 |
| PabiesGI_F2_9_1470 | 0.38120104 |
| X0_15639_01.392 | 0.1285347 |
| PabiesCol1_660 | 0.10539846 |
| PGWD1.1212 | 0.0987013 |
| X2_9845_01.Paab_282 | 0.28238342 |
| CL3507Contig1_03.Paab_191 | 0.11825193 |
| PabiesHB3_5700 | 0.19948187 |
| UMN_853_01.Paab_38 | 0.32891247 |
| PabiesPRR1_1632 | 0.22797927 |
| PGWD1.0794 | 0.14615385 |
| PaPHYO_RII283 | 0.21649485 |
| PGLM2.0818 | 0.45108696 |
| PGLM2.0798 | 0.27272727 |
| PabiesCry_454 | 0.2159383 |
| PabiesPRR1_1039 | 0.34748011 |
| X2_5636_01.Paab_209 | 0.4 |
| X0_17215_01.Paab_225 | 0.49206349 |
| X2_9280_01.Paab_123 | 0.44507042 |
| PGWD1.0344 | 0.47668394 |
| PGWD1.1282 | 0.15762274 |
| X0_7171_01.Paab_233 | 0.26165803 |
| X0_2354_01.Paab_194 | 0.16494845 |
| X0_11772_01.Paab_103 | 0.45876289 |
| PaFTL2pr_1560 | 0.37886598 |

| | |
|---|---|
| UMN_4748_01.Paab_38 | 0.25950064 |
| PabiesHB3_2316 | 0.46085575 |
| PGWD1.0909 | 0.36488568 |
| X0_11090_01.Paab_251 | 0.21191202 |
| PabiesGI_F8_56_324 | 0.26502591 |
| PGLM2.1528 | 0.37333704 |
| PaPHYO_RIV211 | 0.10724846 |
| PGLM2.1477 | 0.19975762 |
| CL3771Contig1_04.Paab_419 | 0.13139584 |
| PabiesGI_F2_9_1470 | 0.38686042 |
| X0_15639_01.392 | 0.13834044 |
| PabiesCol1_660 | 0.10010137 |
| PGWD1.1212 | 0.10340909 |
| X2_9845_01.Paab_282 | 0.27560393 |
| CL3507Contig1_03.Paab_191 | 0.11607876 |
| PabiesHB3_5700 | 0.20050804 |
| UMN_853_01.Paab_38 | 0.35309555 |
| PabiesPRR1_1632 | 0.24127582 |
| PGWD1.0794 | 0.14887285 |
| PaPHYO_RII283 | 0.21347132 |
| PGLM2.0818 | 0.45154603 |
| PGLM2.0798 | 0.28992492 |
| PabiesCry_454 | 0.22076167 |
| PabiesPRR1_1039 | 0.35740236 |
| X2_5636_01.Paab_209 | 0.3739893 |
| X0_17215_01.Paab_225 | 0.49513705 |
| X2_9280_01.Paab_123 | 0.45586059 |
| PGWD1.0344 | 0.46898594 |
| PGWD1.1282 | 0.16706497 |
| X0_7171_01.Paab_233 | 0.27560393 |
| X0_2354_01.Paab_194 | 0.16456219 |
| X0_11772_01.Paab_103 | 0.47998082 |
| PaFTL2pr_1560 | 0.36135419 |

| | |
|---|---|
| X0_7921_01.Paab_212 | 0.02417 |
| PaKN4b_01687n | 0.02527 |
| PaKN1b_08525f | 0.02557 |
| PGWD1.0042 | 0.02677 |
| PabiesPRR1_2381 | 0.02678 |
| PGLM2.0944 | 0.02695 |
| PGWD1.0158 | 0.02696 |
| PabiesPRR1_1039 | 0.02776 |
| X0_489_01.Paab_316 | 0.02843 |
| PabiesPRR3_F2_331 | 0.03200 |
| X0_10267_01.Paab_148 | 0.03273 |
| CL3582Contig1_03.Paab_63 | 0.03331 |
| PabiesFT1_1718 | 0.03521 |
| PGLM2.0828 | 0.03729 |
| PGWD1.0152 | 0.03757 |
| CL1758Contig1_04.288 | 0.03794 |
| PabiesFT1_1050 | 0.03823 |
| PGLM2.0391 | 0.03874 |
| PGLM2.0081 | 0.03875 |
| PGWD1.0220 | 0.03963 |
| PGLM2.1514 | 0.03999 |
| PGWD1.1223 | 0.04005 |
| X0_10515_01.Paab_158 | 0.04066 |
| PGLM2.0624 | 0.04417 |
| X0_11772_01.Paab_103 | 0.04421 |
| PGWD1.1212 | 0.04553 |
| PGWD1.0362 | 0.04603 |
| CL3771Contig1_04.Paab_419 | 0.04619 |
| PGWD1.1295 | 0.04935 |
| X2_6491_01.Paab_360 | 0.05035 |
| X0_7171_01.Paab_233 | 0.05060 |
| PGLM2.0098 | 0.05155 |
| PabiesPRR3_F2_481 | 0.05167 |

| | |
|---|---|
| PabiesMYB2_3403 | 0.16710875 |
| PabiesFT4pr_2046 | 0.49473684 |
| PaMFTL1_2215 | 0.44827586 |
| X2_3867_02.Paab_440 | 0.12403101 |
| X2_9280_01.Paab_338 | 0.47229551 |
| X0_10267_01.Paab_148 | 0.1761658 |
| PGWD1.0418 | 0.18766067 |
| PGWD1.1219 | 0.44356955 |
| X2_4976_01.Paab_176 | 0.2845953 |
| PGLM2.0395 | 0.44675325 |
| PabiesCol1_711 | 0.48924731 |
| PGLM2.0571 | 0.4961039 |
| PGWD1.1304 | 0.51302083 |
| PabiesPRR3_F1_2570 | 0.48806366 |
| PGLM2.0158 | 0.28940568 |
| X2_7725_01.Paab_466 | 0.52105263 |
| X0_17215_01.Paab_108 | 0.44845361 |
| X2_7803_01.Paab_235 | 0.48076923 |
| X0_13957_02.Paab_27 | 0.4947644 |
| PaPHYP_RII177 | 0.2 |
| PabiesPRR1_2381 | 0.23896104 |
| PGWD1.0152 | 0.48 |
| PabiesPrMYB2_932 | 0.39417989 |
| PaKN1b_08525f | 0.11311054 |
| PabiesPRR1_1168 | 0.42447917 |
| PGLM2.0404 | 0.49597855 |
| PGLM2.0288 | 0.28316327 |
| PGWD1.0510 | 0.1038961 |
| X0_7171_01.Paab_359 | 0.2372449 |
| PGLM2.1514 | 0.2408377 |
| X0_10515_01.Paab_158 | 0.33684211 |
| X2_9665_01.Paab_175 | 0.44473684 |
| PGWD1.1518 | 0.31213873 |

| | |
|---|---|
| PabiesMYB2_3403 | 0.1964452 |
| PabiesFT4pr_2046 | 0.48126649 |
| PaMFTL1_2215 | 0.45386661 |
| X2_3867_02.Paab_440 | 0.13458114 |
| X2_9280_01.Paab_338 | 0.45596529 |
| X0_10267_01.Paab_148 | 0.18212772 |
| PGWD1.0418 | 0.19109599 |
| PGWD1.1219 | 0.43554704 |
| X2_4976_01.Paab_176 | 0.30116673 |
| PGLM2.0395 | 0.43214286 |
| PabiesCol1_711 | 0.49989856 |
| PGLM2.0571 | 0.49475446 |
| PGWD1.1304 | 0.49980622 |
| PabiesPRR3_F1_2570 | 0.49836334 |
| PGLM2.0158 | 0.25909413 |
| X2_7725_01.Paab_466 | 0.50126371 |
| X0_17215_01.Paab_108 | 0.43602653 |
| X2_7803_01.Paab_235 | 0.48729679 |
| X0_13957_02.Paab_27 | 0.5013089 |
| PaPHYP_RII177 | 0.19301486 |
| PabiesPRR1_2381 | 0.24553571 |
| PGWD1.0152 | 0.49873797 |
| PabiesPrMYB2_932 | 0.42132612 |
| PaKN1b_08525f | 0.11607876 |
| PabiesPRR1_1168 | 0.42784147 |
| PGLM2.0404 | 0.47469299 |
| PGLM2.0288 | 0.2990207 |
| PGWD1.0510 | 0.10340909 |
| X0_7171_01.Paab_359 | 0.24001448 |
| PGLM2.1514 | 0.25086916 |
| X0_10515_01.Paab_158 | 0.3511179 |
| X2_9665_01.Paab_175 | 0.43990765 |
| PGWD1.1518 | 0.46599648 |

| | |
|---|---|
| PabiesPRR1_240 | 0.05188 |
| CL3771Contig1_04.Paab_68 | 0.05303 |
| PGLM2.0288 | 0.05303 |
| PabiesHB3_385 | 0.05327 |
| PGLM2.0901 | 0.05419 |
| X2_4976_01.Paab_176 | 0.05502 |
| PabiesPRR1_1632 | 0.05511 |
| PGWD1.1282 | 0.05652 |
| PGWD1.0184 | 0.05676 |
| PGLM2.0285 | 0.05867 |
| X0_2433_01.Paab_290 | 0.05902 |
| PGLM2.0798 | 0.05932 |
| PabiesPRR1_3828 | 0.06173 |
| PGWD1.1492 | 0.06371 |
| PGWD1.1501 | 0.06392 |
| PabiesPrMYB2_932 | 0.06443 |
| X2_4892_01.Paab_39 | 0.06665 |
| UMN_853_01.Paab_38 | 0.06849 |
| CL1148Contig1_08.Paab_134 | 0.07018 |
| X0_177_01.Paab_165 | 0.07086 |
| X0_15639_01.392 | 0.07088 |
| PGWD1.0589 | 0.07260 |
| X2_3947_01.Paab_298 | 0.07381 |
| PGWD1.0909 | 0.07654 |
| CL304Contig1_01.202 | 0.07676 |
| X2_3867_02.Paab_440 | 0.07839 |
| PGLM2.0565 | 0.08150 |
| PGLM2.1528 | 0.08366 |
| PaPHYO_RIV211 | 0.08681 |
| PaMFTL1_1613 | 0.08708 |
| PabiesPRR1_3883 | 0.08722 |
| PabiesZTL_367 | 0.09159 |
| PGLM2.0021 | 0.09986 |

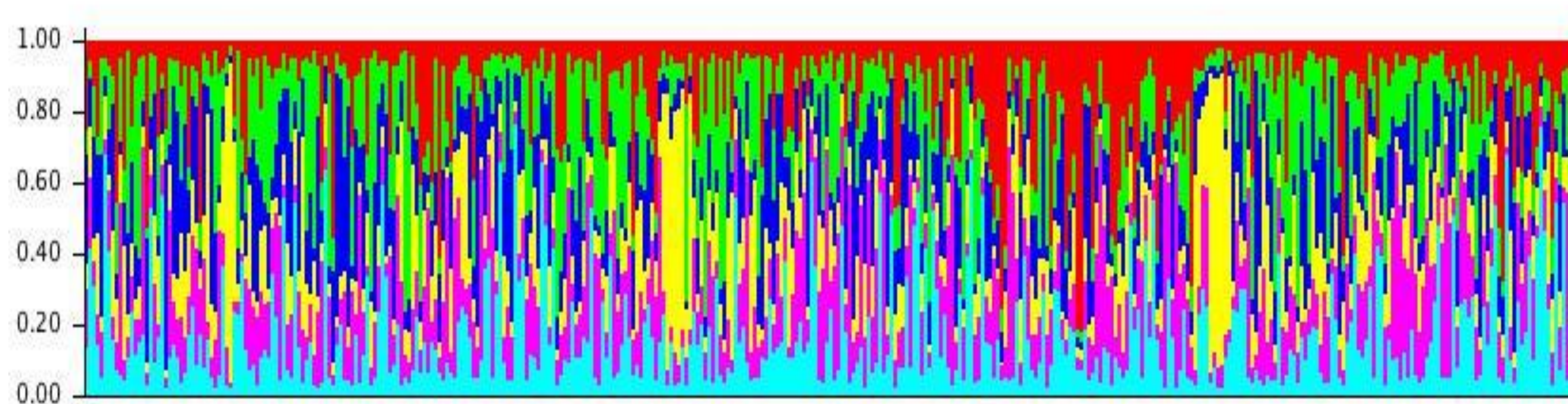
| | |
|---|---|
| PGLM2.0592 | 0.50130548 |
| X0_13978_01.Paab_102 | 0.1974026 |
| PGLM2.0624 | 0.47354497 |
| X0_10754_01.Paab_320 | 0.2519685 |
| PabiesFT1_911 | 0.31524548 |
| PGLM2.0154 | 0.16537468 |
| PGLM2.1091 | 0.09768638 |
| PGLM2.0610 | 0.27105263 |
| PGWD1.1070 | 0.46276596 |
| PGWD1.1295 | 0.47058824 |
| X0_9383_01.Paab_438 | 0.14136126 |
| PGLM2.0944 | 0.28981723 |
| PabiesPRR1_240 | 0.38522427 |
| PGLM2.0828 | 0.22997416 |
| PGWD1.0118 | 0.14507772 |
| **Mean Ho** | **0.3262** |

| | |
|---|---|
| PGLM2.0592 | 0.49712247 |
| X0_13978_01.Paab_102 | 0.19080087 |
| PGLM2.0624 | 0.49542826 |
| X0_10754_01.Paab_320 | 0.28719436 |
| PabiesFT1_911 | 0.31958335 |
| PGLM2.0154 | 0.15640439 |
| PGLM2.1091 | 0.09779238 |
| PGLM2.0610 | 0.2720143 |
| PGWD1.1070 | 0.45685106 |
| PGWD1.1295 | 0.49501799 |
| X0_9383_01.Paab_438 | 0.13171456 |
| PGLM2.0944 | 0.29784493 |
| PabiesPRR1_240 | 0.40630453 |
| PGLM2.0828 | 0.23888086 |
| PGWD1.0118 | 0.14373192 |
| **Mean He** | **0.3258** |

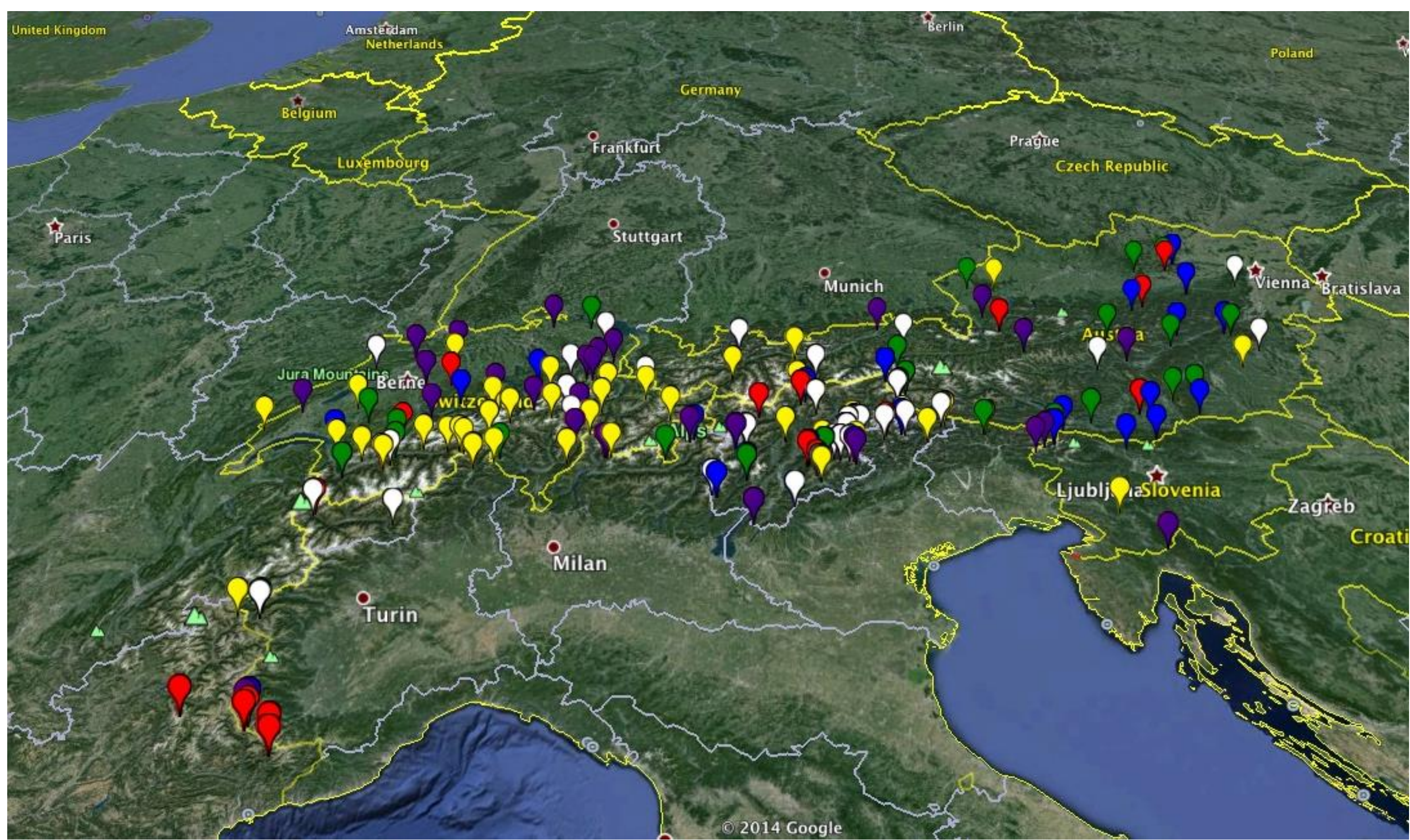
| | |
|---|---|
| PGLM2.0923 | 0.11785 |
| X0_10754_01.Paab_320 | 0.12266 |
| PGWD1.0634 | 0.12295 |
| PaPHYO_RI145 | 0.12756 |
| PGLM2.0584 | 0.13083 |
| CL1148Contig1_08.Paab_225 | 0.14295 |
| PabiesMYB2_3403 | 0.14934 |
| PGLM2.1269 | 0.16430 |
| PGLM2.0134 | 0.16593 |
| UMN_3055_01.Paab_224 | 0.19264 |
| PGWD1.0853 | 0.20557 |
| PGLM2.1477 | 0.20957 |
| X0_8531_01.Paab_363 | 0.30529 |
| PGWD1.1518 | 0.33017 |
| PGWD1.1034 | 0.46765 |
| **Mean Fis** | **4.08E-03** |

Supplement 6. (A) Barplot when K=6 for original order for 394 individuals assuming existence of panmixia. Each individual is represented by a single vertical line, which is partitioned into K colored segments that represent that individual's estimated membership fraction in each of K inferred clusters. (B) Individuals plotted on Google map when K=6 indicating process such as mating and individuals movement are uniform (population in panmixia).

(A)



(B)



Supplement 7. Fst coefficient estimated as the posterior mean using model averaging (BayeScan)

| SNP | prob | log10(PO) | qval | alpha | Fst |
|---|---|---|---|---|---|
| 1 | 0.10082 | -0.950299 | 0.846194 | 0.057304 | 0.014897 |
| 2 | 0.079816 | -1.0618 | 0.88903 | -0.035053 | 0.012822 |
| 3 | 0.065613 | -1.1535 | 0.90316 | -0.028613 | 0.012881 |
| 4 | 0.30206 | -0.36372 | 0.36972 | -0.427 | 0.010312 |
| 5 | 0.071614 | -1.1127 | 0.89798 | -0.028071 | 0.012928 |
| 6 | 0.095019 | -0.97883 | 0.86141 | -0.06329 | 0.012594 |
| 7 | 0.086017 | -1.0264 | 0.88081 | 0.035907 | 0.014068 |
| 8 | 0.080616 | -1.0571 | 0.88747 | -0.016852 | 0.013305 |
| 9 | 0.087818 | -1.0165 | 0.8775 | 0.015633 | 0.01393 |
| 10 | 0.065613 | -1.1535 | 0.90316 | 0.0014538 | 0.01323 |
| 11 | 0.070614 | -1.1193 | 0.89931 | -0.023277 | 0.012932 |
| 12 | 0.075615 | -1.0872 | 0.8944 | -0.041188 | 0.012757 |
| 13 | 0.093019 | -0.98903 | 0.86605 | -0.056935 | 0.012688 |
| 14 | 0.09962 | -0.95608 | 0.84802 | 0.059047 | 0.014345 |
| 15 | 0.083017 | -1.0432 | 0.88422 | -0.01475 | 0.013203 |
| 16 | 0.09762 | -0.96585 | 0.85443 | -0.068909 | 0.012594 |
| 17 | 0.088218 | -1.0143 | 0.87495 | -0.0027569 | 0.013682 |
| 18 | 0.10722 | -0.92046 | 0.82292 | 0.062538 | 0.014645 |
| 19 | 0.10202 | -0.94458 | 0.83884 | -0.070137 | 0.012585 |
| 20 | 0.096219 | -0.9728 | 0.85844 | 0.015005 | 0.014213 |
| 21 | 0.093019 | -0.98903 | 0.86605 | -0.063531 | 0.012563 |
| 22 | 0.064013 | -1.165 | 0.90387 | -0.0093525 | 0.013056 |
| 23 | 0.65513 | 0.27867 | 0.26032 | 0.8337 | 0.03567 |
| 24 | 0.079616 | -1.063 | 0.88977 | -0.030404 | 0.012882 |
| 25 | 0.077816 | -1.0738 | 0.89185 | -0.028228 | 0.013001 |
| 26 | 0.076615 | -1.0811 | 0.89315 | 0.011553 | 0.013493 |
| 27 | 0.85197 | 0.76008 | 0.14803 | 1.1283 | 0.044796 |
| 28 | 0.13003 | -0.82548 | 0.71396 | 0.10465 | 0.01652 |
| 29 | 0.080216 | -1.0594 | 0.88846 | 0.0039764 | 0.013464 |
| 30 | 0.11362 | -0.89215 | 0.80295 | -0.081506 | 0.012477 |
| 31 | 0.070814 | -1.118 | 0.89892 | -0.0029809 | 0.013218 |
| 32 | 0.087818 | -1.0165 | 0.8775 | -0.028485 | 0.013113 |
| 33 | 0.087217 | -1.0198 | 0.87921 | -0.020272 | 0.013177 |
| 34 | 0.082016 | -1.0489 | 0.88513 | 0.014476 | 0.013835 |
| 35 | 0.11582 | -0.88274 | 0.78635 | -0.094102 | 0.01241 |
| 36 | 0.09822 | -0.9629 | 0.85139 | -0.071333 | 0.012564 |
| 37 | 0.089818 | -1.0058 | 0.8713 | 0.013322 | 0.013968 |
| 38 | 0.090218 | -1.0036 | 0.86967 | -0.042268 | 0.012913 |
| 39 | 0.079616 | -1.063 | 0.88977 | -0.017874 | 0.013087 |
| 40 | 0.12783 | -0.83399 | 0.73373 | -0.10683 | 0.012299 |
| 41 | 0.062813 | -1.1738 | 0.9047 | -0.010433 | 0.01306 |
| 42 | 0.11282 | -0.89561 | 0.8106 | 0.06993 | 0.014672 |
| 43 | 0.081816 | -1.0501 | 0.88535 | -0.048654 | 0.01271 |
| 44 | 0.079216 | -1.0653 | 0.89031 | -0.031914 | 0.012961 |
| 45 | 0.082016 | -1.0489 | 0.88513 | 0.029261 | 0.013801 |
| 46 | 0.064613 | -1.1607 | 0.90363 | -0.010742 | 0.013067 |
| 47 | 0.083017 | -1.0432 | 0.88422 | 0.032612 | 0.013844 |
| 48 | 0.079016 | -1.0665 | 0.89083 | -0.044301 | 0.012766 |
| 49 | 0.09922 | -0.95802 | 0.84975 | 0.051549 | 0.014276 |

| 50 | 0.091818 | -0.99524 | 0.867 | -0.05791 | 0.012628 |
|---|---|---|---|---|---|
| 51 | 0.089818 | -1.0058 | 0.8713 | -0.014362 | 0.013366 |
| 52 | 0.10542 | -0.92869 | 0.83238 | 0.067294 | 0.014782 |
| 53 | 0.070814 | -1.118 | 0.89892 | -0.0012604 | 0.013262 |
| 54 | 0.066013 | -1.1507 | 0.9028 | -0.0095085 | 0.013092 |
| 55 | 0.092218 | -0.99316 | 0.86653 | -0.055311 | 0.012666 |
| 56 | 0.079216 | -1.0653 | 0.89031 | 0.010718 | 0.013533 |
| 57 | 0.068814 | -1.1314 | 0.90096 | -0.016336 | 0.013043 |
| 58 | 0.081416 | -1.0524 | 0.88664 | -0.043007 | 0.012739 |
| 59 | 0.080216 | -1.0594 | 0.88846 | -0.043399 | 0.012758 |
| 60 | 0.063413 | -1.1694 | 0.90411 | 0.0046024 | 0.013247 |
| 61 | 0.097419 | -0.96684 | 0.85514 | -0.07341 | 0.012481 |
| 62 | 0.065413 | -1.155 | 0.90327 | 0.023726 | 0.01354 |
| 63 | 0.078216 | -1.0713 | 0.89168 | -0.021643 | 0.013035 |
| 64 | 0.072214 | -1.1088 | 0.89771 | -0.026862 | 0.01291 |
| 65 | 0.09922 | -0.95802 | 0.84975 | -0.070396 | 0.012565 |
| 66 | 0.064813 | -1.1592 | 0.90351 | 0.00014235 | 0.013237 |
| 67 | 0.070614 | -1.1193 | 0.89931 | -0.0013348 | 0.01323 |
| 68 | 0.069014 | -1.13 | 0.90059 | -0.0040371 | 0.013223 |
| 69 | 0.10382 | -0.93611 | 0.83509 | -0.078797 | 0.01245 |
| 70 | 0.11442 | -0.88871 | 0.7937 | 0.070042 | 0.015645 |
| 71 | 0.066013 | -1.1507 | 0.9028 | -0.0035155 | 0.01313 |
| 72 | 0.069414 | -1.1273 | 0.90046 | -0.0028567 | 0.01321 |
| 73 | 0.11282 | -0.89561 | 0.8106 | 0.065525 | 0.014652 |
| 74 | 0.068614 | -1.1327 | 0.90109 | 0.012372 | 0.013427 |
| 75 | 0.12202 | -0.85704 | 0.76793 | -0.10786 | 0.012261 |
| 76 | 0.13043 | -0.82394 | 0.70196 | -0.11546 | 0.012206 |
| 77 | 0.087618 | -1.0176 | 0.87779 | 0.029723 | 0.013903 |
| 78 | 0.11022 | -0.90701 | 0.81511 | -0.092376 | 0.012392 |
| 79 | 0.069414 | -1.1273 | 0.90046 | 0.0019057 | 0.013225 |
| 80 | 0.085017 | -1.0319 | 0.88183 | -0.046421 | 0.012722 |
| 81 | 0.095019 | -0.97883 | 0.86141 | -0.066277 | 0.01258 |
| 82 | 0.093419 | -0.98697 | 0.86459 | -0.069295 | 0.012536 |
| 83 | 0.15143 | -0.74847 | 0.56146 | 0.12023 | 0.015912 |
| 84 | 0.10862 | -0.91415 | 0.82108 | -0.090969 | 0.012375 |
| 85 | 0.10702 | -0.92137 | 0.82634 | -0.077278 | 0.01253 |
| 86 | 0.087417 | -1.0187 | 0.87837 | -0.0070443 | 0.013521 |
| 87 | 0.071414 | -1.114 | 0.89825 | -0.020727 | 0.012986 |
| 88 | 0.081416 | -1.0524 | 0.88664 | 0.032915 | 0.013786 |
| 89 | 0.13223 | -0.81709 | 0.65189 | -0.12519 | 0.012151 |
| 90 | 0.075015 | -1.091 | 0.89529 | -0.017402 | 0.013021 |
| 91 | 0.090618 | -1.0015 | 0.86836 | -0.05655 | 0.012639 |
| 92 | 0.061612 | -1.1827 | 0.90518 | -0.010627 | 0.01304 |
| 93 | 0.087818 | -1.0165 | 0.8775 | -0.033803 | 0.012977 |
| 94 | 0.12382 | -0.84978 | 0.74211 | -0.10377 | 0.012306 |
| 95 | 0.068014 | -1.1368 | 0.90133 | 0.0017037 | 0.013282 |
| 96 | 0.13103 | -0.82165 | 0.68799 | 0.1034 | 0.01662 |
| 97 | 0.078216 | -1.0713 | 0.89168 | -0.036852 | 0.012777 |
| 98 | 0.084817 | -1.033 | 0.88233 | -0.00063773 | 0.013604 |
| 99 | 0.09882 | -0.95997 | 0.85058 | -0.064014 | 0.012625 |
| 100 | 0.081616 | -1.0512 | 0.88579 | -0.020255 | 0.013063 |
| 101 | 0.076215 | -1.0835 | 0.89378 | -0.0077562 | 0.013145 |

| | | | | | |
|---|---|---|---|---|---|
| 102 | 0.063613 | -1.1679 | 0.90399 | -0.010297 | 0.013041 |
| 103 | 0.12783 | -0.83399 | 0.73373 | -0.11593 | 0.012222 |
| 104 | 0.080416 | -1.0582 | 0.88767 | -0.042695 | 0.012767 |
| 105 | 0.15463 | -0.73775 | 0.5136 | 0.12908 | 0.016547 |
| 106 | 0.070414 | -1.1206 | 0.89944 | -0.010035 | 0.013117 |
| 107 | 0.11402 | -0.89043 | 0.797 | -0.1031 | 0.012333 |
| 108 | 0.067413 | -1.1409 | 0.9017 | -0.024483 | 0.012916 |
| 109 | 0.075215 | -1.0897 | 0.89485 | -0.029479 | 0.012886 |
| 110 | 0.094419 | -0.98187 | 0.86196 | -0.072212 | 0.012554 |
| 111 | 0.083017 | -1.0432 | 0.88422 | -0.041755 | 0.012775 |
| 112 | 0.086417 | -1.0241 | 0.88055 | 0.02933 | 0.013916 |
| 113 | 0.084417 | -1.0353 | 0.88328 | -0.041627 | 0.012754 |
| 114 | 0.078816 | -1.0677 | 0.891 | -0.039454 | 0.012794 |
| 115 | 0.089818 | -1.0058 | 0.8713 | -0.065192 | 0.012593 |
| 116 | 0.091218 | -0.99838 | 0.86791 | -0.066629 | 0.012545 |
| 117 | 0.070214 | -1.122 | 0.89957 | 0.0091001 | 0.013386 |
| 118 | 0.091418 | -0.99733 | 0.86746 | -0.062508 | 0.012639 |
| 119 | 0.084417 | -1.0353 | 0.88328 | -0.0090183 | 0.013435 |
| 120 | 0.11582 | -0.88274 | 0.78635 | -0.10407 | 0.012283 |
| 121 | 0.075015 | -1.091 | 0.89529 | -0.03072 | 0.012859 |
| 122 | 0.075415 | -1.0885 | 0.8947 | -0.030528 | 0.012879 |
| 123 | 0.072214 | -1.1088 | 0.89771 | -0.0019451 | 0.01323 |
| 124 | 0.11002 | -0.9079 | 0.81719 | -0.089156 | 0.012398 |
| 125 | 0.084817 | -1.033 | 0.88233 | 0.037983 | 0.013879 |
| 126 | 0.088218 | -1.0143 | 0.87495 | -0.060394 | 0.012603 |
| 127 | 0.062813 | -1.1738 | 0.9047 | -0.017201 | 0.012954 |
| 128 | 0.063213 | -1.1708 | 0.90435 | 0.0030615 | 0.013231 |
| 129 | 0.074215 | -1.096 | 0.89588 | -0.026448 | 0.012916 |
| 130 | 0.10142 | -0.94743 | 0.84325 | -0.079236 | 0.012451 |
| 131 | 0.075815 | -1.086 | 0.89409 | -0.037096 | 0.012839 |
| 132 | 0.061212 | -1.1857 | 0.9053 | 0.0013268 | 0.013217 |
| 133 | 0.065813 | -1.1521 | 0.90292 | -0.0038352 | 0.013159 |
| 134 | 0.084017 | -1.0375 | 0.88352 | -0.032124 | 0.013003 |
| 135 | 0.079016 | -1.0665 | 0.89083 | -0.03857 | 0.012819 |
| 136 | 0.073215 | -1.1024 | 0.89673 | -0.021339 | 0.013002 |
| 137 | 0.068814 | -1.1314 | 0.90096 | -0.012059 | 0.013096 |
| 138 | 0.096619 | -0.97081 | 0.85716 | -0.035635 | 0.013124 |
| 139 | 0.087417 | -1.0187 | 0.87837 | -0.056909 | 0.012663 |
| 140 | 0.074215 | -1.096 | 0.89588 | -0.024545 | 0.01293 |
| 141 | 0.13343 | -0.81256 | 0.6279 | -0.12104 | 0.012193 |
| 142 | 0.069614 | -1.126 | 0.90021 | 0.011314 | 0.013439 |
| 143 | 0.088818 | -1.0111 | 0.87391 | -0.063247 | 0.012629 |
| 144 | 0.087217 | -1.0198 | 0.87921 | -0.051657 | 0.012693 |
| 145 | 0.086617 | -1.023 | 0.88002 | 0.037491 | 0.013936 |
| 146 | 0.087217 | -1.0198 | 0.87921 | 0.03876 | 0.014078 |
| 147 | 0.10582 | -0.92685 | 0.82795 | 0.068398 | 0.014611 |
| 148 | 0.077215 | -1.0774 | 0.89267 | -0.041944 | 0.012735 |
| 149 | 0.087818 | -1.0165 | 0.8775 | -0.038575 | 0.012871 |
| 150 | 0.24265 | -0.49432 | 0.44725 | 0.23683 | 0.018816 |
| 151 | 0.077015 | -1.0786 | 0.89283 | -0.024048 | 0.012957 |
| 152 | 0.079616 | -1.063 | 0.88977 | -0.043848 | 0.01272 |
| 153 | 0.073415 | -1.1011 | 0.89645 | -0.024216 | 0.012951 |
| 154 | 0.070814 | -1.118 | 0.89892 | -0.014438 | 0.013034 |

| | | | | | |
|---|---|---|---|---|---|
| 155 | 0.095019 | -0.97883 | 0.86141 | -0.065047 | 0.012596 |
| 156 | 0.071214 | -1.1153 | 0.89852 | -0.0046079 | 0.013177 |
| 157 | 0.069614 | -1.126 | 0.90021 | 0.015852 | 0.013473 |
| 158 | 0.094019 | -0.9839 | 0.86357 | -0.064073 | 0.012609 |
| 159 | 0.088418 | -1.0133 | 0.87427 | -0.05163 | 0.012708 |
| 160 | 0.096019 | -0.9738 | 0.85905 | -0.059119 | 0.012648 |
| 161 | 0.067013 | -1.1437 | 0.90207 | -0.0088133 | 0.013093 |
| 162 | 0.071214 | -1.1153 | 0.89852 | -0.016787 | 0.013071 |
| 163 | 0.090418 | -1.0026 | 0.86924 | -0.061049 | 0.012592 |
| 164 | 0.062613 | -1.1753 | 0.90482 | 0.0054506 | 0.013273 |
| 165 | 0.078416 | -1.0701 | 0.89134 | 0.0015925 | 0.013304 |
| 166 | 0.093619 | -0.98595 | 0.86409 | -0.063522 | 0.012604 |
| 167 | 0.079616 | -1.063 | 0.88977 | 0.018617 | 0.013547 |
| 168 | 0.11082 | -0.90436 | 0.81291 | -0.084159 | 0.012484 |
| 169 | 0.11782 | -0.87432 | 0.77784 | -0.10779 | 0.012285 |
| 170 | 0.14563 | -0.7684 | 0.59807 | -0.13931 | 0.012048 |
| 171 | 0.067614 | -1.1396 | 0.90158 | 0.0073837 | 0.013291 |
| 172 | 0.085417 | -1.0297 | 0.88107 | 0.0312 | 0.013858 |
| 173 | 0.066013 | -1.1507 | 0.9028 | -0.014082 | 0.013042 |
| 174 | 0.09782 | -0.96487 | 0.85296 | 0.021335 | 0.014369 |
| 175 | 0.076415 | -1.0823 | 0.89331 | -0.041744 | 0.012756 |
| 176 | 0.066413 | -1.1479 | 0.90244 | -0.010606 | 0.013087 |
| 177 | 0.085017 | -1.0319 | 0.88183 | -0.020249 | 0.01319 |
| 178 | 0.090018 | -1.0047 | 0.87009 | -0.052983 | 0.012703 |
| 179 | 0.088818 | -1.0111 | 0.87391 | 0.00081756 | 0.013643 |
| 180 | 0.080216 | -1.0594 | 0.88846 | -0.038095 | 0.012824 |
| 181 | 0.073215 | -1.1024 | 0.89673 | -0.022865 | 0.012973 |
| 182 | 0.069814 | -1.1246 | 0.89983 | 0.016466 | 0.013503 |
| 183 | 0.063213 | -1.1708 | 0.90435 | 0.0070222 | 0.0133 |
| 184 | 0.088018 | -1.0154 | 0.87562 | -0.0033285 | 0.013509 |
| 185 | 0.11362 | -0.89215 | 0.80295 | -0.097184 | 0.012358 |
| 186 | 0.081016 | -1.0547 | 0.88706 | 0.018077 | 0.01354 |
| 187 | 0.079016 | -1.0665 | 0.89083 | -0.035444 | 0.012835 |
| 188 | 0.080216 | -1.0594 | 0.88846 | -0.023142 | 0.012967 |
| 189 | 0.11442 | -0.88871 | 0.7937 | -0.09128 | 0.012396 |
| 190 | 0.090418 | -1.0026 | 0.86924 | -0.053887 | 0.012668 |
| 191 | 0.082817 | -1.0443 | 0.88445 | -0.044322 | 0.012748 |
| 192 | 0.10182 | -0.94553 | 0.84 | 0.050624 | 0.014379 |
| 193 | 0.087818 | -1.0165 | 0.8775 | 0.013687 | 0.013761 |
| 194 | 0.12202 | -0.85704 | 0.76793 | -0.10566 | 0.012299 |
| 195 | 0.076215 | -1.0835 | 0.89378 | -0.020427 | 0.013032 |
| 196 | 0.089018 | -1.01 | 0.8732 | -0.053705 | 0.012689 |
| 197 | 0.09762 | -0.96585 | 0.85443 | 0.023042 | 0.014181 |
| 198 | 0.093219 | -0.988 | 0.86509 | -0.03404 | 0.013117 |
| 199 | 0.097219 | -0.96783 | 0.85583 | -0.050957 | 0.012766 |
| 200 | 0.076815 | -1.0798 | 0.89299 | -0.02179 | 0.013021 |
| 201 | 0.71194 | 0.39297 | 0.21804 | 1.0147 | 0.043965 |
| 202 | 0.10082 | -0.9503 | 0.84619 | -0.074771 | 0.012515 |
| 203 | 0.072414 | -1.1075 | 0.89729 | -0.0038399 | 0.013243 |
| 204 | 0.068214 | -1.1354 | 0.90121 | -0.0082674 | 0.013137 |
| 205 | 0.079816 | -1.0618 | 0.88903 | 0.023453 | 0.01366 |
| 206 | 0.074415 | -1.0948 | 0.89544 | -0.028455 | 0.012932 |
| 207 | 0.073615 | -1.0998 | 0.89631 | -0.018828 | 0.013043 |

| | | | | | |
|---|---|---|---|---|---|
| 208 | 0.075415 | -1.0885 | 0.8947 | -0.035002 | 0.01286 |
| 209 | 0.094019 | -0.9839 | 0.86357 | -0.06522 | 0.012599 |
| 210 | 0.074015 | -1.0973 | 0.89616 | -0.029732 | 0.012862 |
| 211 | 0.067013 | -1.1437 | 0.90207 | -0.015437 | 0.013031 |
| 212 | 0.11322 | -0.89388 | 0.80566 | -0.095856 | 0.012373 |
| 213 | 0.084417 | -1.0353 | 0.88328 | -0.031968 | 0.012996 |
| 214 | 0.079816 | -1.0618 | 0.88903 | -0.041902 | 0.012801 |
| 215 | 0.062813 | -1.1738 | 0.9047 | 0.0022691 | 0.01323 |
| 216 | 0.062212 | -1.1782 | 0.90506 | 0.003361 | 0.013248 |
| 217 | 0.10442 | -0.93332 | 0.83376 | -0.073676 | 0.012501 |
| 218 | 0.084417 | -1.0353 | 0.88328 | 0.03184 | 0.01381 |
| 219 | 0.089218 | -1.009 | 0.87283 | -0.0068411 | 0.013588 |
| 220 | 0.12342 | -0.85139 | 0.74958 | -0.11469 | 0.012231 |
| 221 | 0.094219 | -0.98289 | 0.86251 | 0.0011068 | 0.013823 |
| 222 | 0.077415 | -1.0762 | 0.89234 | -0.035014 | 0.012842 |
| 223 | 0.077415 | -1.0762 | 0.89234 | 0.033069 | 0.013821 |
| 224 | 0.067213 | -1.1423 | 0.90183 | -0.0070271 | 0.013141 |
| 225 | 0.079216 | -1.0653 | 0.89031 | -0.017918 | 0.013096 |
| 226 | 0.088018 | -1.0154 | 0.87562 | -0.014548 | 0.013282 |
| 227 | 0.070614 | -1.1193 | 0.89931 | -0.025147 | 0.01293 |
| 228 | 0.081416 | -1.0524 | 0.88664 | -0.040387 | 0.012797 |
| 229 | 0.081416 | -1.0524 | 0.88664 | -0.036549 | 0.012821 |
| 230 | 0.086817 | -1.022 | 0.87949 | -0.057679 | 0.012664 |
| 231 | 0.072214 | -1.1088 | 0.89771 | -0.010091 | 0.013102 |
| 232 | 0.072014 | -1.1101 | 0.89784 | 0.004488 | 0.013354 |
| 233 | 0.076215 | -1.0835 | 0.89378 | -0.0085716 | 0.013215 |
| 234 | 0.095219 | -0.97782 | 0.85966 | 0.020628 | 0.014392 |
| 235 | 0.10082 | -0.9503 | 0.84619 | -0.070718 | 0.01255 |
| 236 | 0.065013 | -1.1578 | 0.90339 | -0.010019 | 0.013057 |
| 237 | 0.069614 | -1.126 | 0.90021 | -0.018496 | 0.013011 |
| 238 | 0.071414 | -1.114 | 0.89825 | -0.034537 | 0.012804 |
| 239 | 0.096419 | -0.9718 | 0.85781 | 0.033865 | 0.014053 |
| 240 | 0.10562 | -0.92777 | 0.83097 | -0.081658 | 0.012475 |
| 241 | 0.064413 | -1.1621 | 0.90375 | 0.010117 | 0.013325 |
| 242 | 0.10142 | -0.94743 | 0.84325 | -0.076683 | 0.012506 |
| 243 | 0.081216 | -1.0536 | 0.88685 | -0.02075 | 0.013068 |
| 244 | 0.069814 | -1.1246 | 0.89983 | -0.015899 | 0.013012 |
| 245 | 0.072815 | -1.1049 | 0.89701 | -0.030581 | 0.012867 |
| 246 | 0.060012 | -1.1949 | 0.90554 | -0.015298 | 0.012977 |
| 247 | 0.074215 | -1.096 | 0.89588 | -0.0093922 | 0.013131 |
| 248 | 0.066813 | -1.1451 | 0.90219 | -0.0093158 | 0.013116 |
| 249 | 0.10702 | -0.92137 | 0.82634 | -0.089447 | 0.012409 |
| 250 | 0.060012 | -1.1949 | 0.90554 | -0.0024698 | 0.013125 |
| 251 | 0.10162 | -0.94648 | 0.84112 | -0.077772 | 0.012491 |
| 252 | 0.085017 | -1.0319 | 0.88183 | -0.015346 | 0.013288 |
| 253 | 0.10302 | -0.93986 | 0.83638 | -0.075028 | 0.01254 |
| 254 | 0.077616 | -1.075 | 0.89201 | -0.032046 | 0.01288 |
| 255 | 0.074015 | -1.0973 | 0.89616 | 0.0096781 | 0.01354 |
| 256 | 0.077215 | -1.0774 | 0.89267 | -0.021655 | 0.013007 |
| 257 | 0.13203 | -0.81785 | 0.67153 | -0.12713 | 0.012158 |
| 258 | 0.066413 | -1.1479 | 0.90244 | 0.0034851 | 0.013223 |
| 259 | 0.073015 | -1.1037 | 0.89687 | -0.016141 | 0.013058 |
| 260 | 0.075015 | -1.091 | 0.89529 | -0.0049445 | 0.013293 |

| | | | | | |
|---|---|---|---|---|---|
| 261 | 0.075615 | -1.0872 | 0.8944 | -0.035549 | 0.012819 |
| 262 | 0.09782 | -0.96487 | 0.85296 | -0.070354 | 0.012587 |
| 263 | 0.082416 | -1.0466 | 0.88468 | -0.050849 | 0.012696 |
| 264 | 0.067814 | -1.1382 | 0.90146 | 0.0032176 | 0.013263 |
| 265 | 0.10022 | -0.95318 | 0.84712 | -0.078191 | 0.012502 |
| 266 | 0.097019 | -0.96882 | 0.8565 | -0.064375 | 0.012595 |
| 267 | 0.080616 | -1.0571 | 0.88747 | -0.043415 | 0.012774 |
| 268 | 0.089418 | -1.0079 | 0.87208 | 0.039803 | 0.013989 |
| 269 | 0.087818 | -1.0165 | 0.8775 | -0.059116 | 0.012609 |
| 270 | 0.086417 | -1.0241 | 0.88055 | -0.055548 | 0.012635 |
| 271 | 0.11822 | -0.87265 | 0.7731 | -0.10488 | 0.012317 |
| 272 | 0.078416 | -1.0701 | 0.89134 | -0.029514 | 0.012928 |
| 273 | 0.068814 | -1.1314 | 0.90096 | -0.02051 | 0.012979 |
| 274 | 0.089418 | -1.0079 | 0.87208 | -0.05537 | 0.012664 |
| 275 | 0.10562 | -0.92777 | 0.83097 | 0.067551 | 0.014541 |
| 276 | 0.10222 | -0.94363 | 0.83763 | -0.078929 | 0.012468 |
| 277 | 0.062212 | -1.1782 | 0.90506 | 0.00012207 | 0.013202 |
| 278 | 0.072615 | -1.1062 | 0.89715 | 0.010363 | 0.013395 |
| 279 | 0.081616 | -1.0512 | 0.88579 | -0.007806 | 0.013322 |
| 280 | 0.12202 | -0.85704 | 0.76793 | 0.084107 | 0.014891 |
| 281 | 0.089218 | -1.009 | 0.87283 | -0.057623 | 0.012642 |
| 282 | 0.086617 | -1.023 | 0.88002 | 0.0032196 | 0.013608 |
| 283 | 0.076015 | -1.0848 | 0.89394 | 0.024764 | 0.013639 |
| 284 | 0.10902 | -0.91235 | 0.81918 | -0.08205 | 0.012476 |

Supplement 8: Samβada results for univariate models. List of loci detected under the BEST model when threshold of 0.01 used indicating measures of regression (Efron, McFadden, McFaddenAdj, CoxSnell, Nagelkerke), Akaike information criterion (AIC), Bayesian information criterion (BIC), Parameters β for regression (Batea_0 and Beta_1).

| Marker | Env_1 | Loglikelihood | Gscore | WaldScore | NumError | Efron | McFadden | McFaddenAdj | CoxSnell | Nagelkerke | AIC | BIC | Beta_0 | Beta_1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 160_BB | ppt_win | -224.4376832 | 28.24125588 | 26.19277394 | 0 | 0.098208502 | 0.059191515 | 0.050807821 | 0.072918522 | 0.030327029 | 452.8753663 | 472.56168 | -1.911515899 | 0.015787804 |
| 7_AA | ppt_sum | -104.8122197 | 19.18984068 | 16.74221474 | 0 | 0.075780996 | 0.083866447 | 0.066385021 | 0.049372347 | 0.062654891 | 213.6244393 | 233.3745842 | 0.290170755 | -0.024202408 |
| 160_AA | AI_spring | -175.7283632 | 17.36421137 | 16.65538663 | 0 | 0.058371737 | 0.04708033 | 0.036234958 | 0.045485883 | 0.028230772 | 355.4567265 | 375.1430402 | -1.96269716 | 0.534087802 |
| 25_BB | AI_spring | -254.4247472 | 14.78875768 | 14.08623537 | 0 | 0.075423312 | 0.028242317 | 0.020603456 | 0.03836825 | 0.013317592 | 512.8494944 | 532.5890712 | 0.406946573 | -0.401797977 |
| 160_AA | ppt_win | -176.421083 | 15.97877185 | 14.07297719 | 0 | 0.051843497 | 0.043323928 | 0.032478557 | 0.041933924 | 0.025929762 | 356.842166 | 376.5284797 | -0.362207754 | -0.015221622 |
| 183_BB | AI_spring | -257.44203 | 14.29546704 | 13.55672464 | 0 | 0.050788612 | 0.027014399 | 0.019455528 | 0.036173507 | 0.012891627 | 518.88406 | 538.7280814 | -0.021398937 | 0.394899121 |
| 174_AA | t_win | -68.30816205 | 14.76476205 | 13.08827444 | 0 | 0.045559385 | 0.09753373 | 0.07111035 | 0.03762398 | 0.08115835 | 140.6163241 | 160.4292974 | -3.071267494 | -0.421571509 |
| 278_BB | ppt_aut | -159.7674513 | 15.33705443 | 12.65257429 | 0 | 0.066100714 | 0.045799757 | 0.033854893 | 0.03856256 | 0.029492912 | 323.5349027 | 343.3994896 | -0.123523187 | 0.018281213 |
| 278_AB | ppt_aut | -154.6469324 | 15.16984895 | 12.49290086 | 0 | 0.040883995 | 0.046753607 | 0.034425572 | 0.038150272 | 0.030561732 | 313.2938647 | 333.1584517 | 0.0957661 | -0.018653562 |
| 53_AA | ppt_win | -136.108697 | 14.52291671 | 12.40240165 | 0 | 0.089961772 | 0.050648326 | 0.036698421 | 0.036645656 | 0.034901596 | 276.2173939 | 296.0717113 | 0.771947745 | 0.018030778 |
| 186_BB | ppt_win | -77.39457786 | 12.67304761 | 12.29611383 | 0 | 0.04756635 | 0.075677062 | 0.051791075 | 0.033230128 | 0.061059067 | 158.7891557 | 178.4968598 | -4.552283128 | 0.020235258 |
| 53_AB | ppt_win | -134.2423974 | 14.26183921 | 12.17458731 | 0 | 0.039873694 | 0.05044035 | 0.036293409 | 0.035998883 | 0.034947054 | 272.4847949 | 292.3391123 | -0.796566141 | -0.018042827 |
| 204_BB | altitude | -249.009191 | 12.95805333 | 12.16825325 | 0 | 0.061078611 | 0.025359395 | 0.017531246 | 0.033875825 | 0.01212343 | 502.018382 | 521.7367385 | -1.790721317 | 0.001100956 |
| 165_BB | ppt_win | -175.5579592 | 12.26429668 | 12.10752549 | 0 | 0.043066348 | 0.03375059 | 0.022742836 | 0.032865499 | 0.020088845 | 355.1159184 | 374.7373658 | -2.369950247 | 0.01213414 |
| 86_AB | ppt_win | -142.2635915 | 13.90717299 | 12.00737781 | 0 | 0.041799013 | 0.046600442 | 0.03319716 | 0.035119565 | 0.03154902 | 288.527183 | 308.3815004 | -0.763816568 | -0.017008948 |
| 99_AA | ppt_sum | -258.6960257 | 12.22265718 | 11.81241584 | 0 | 0.048984086 | 0.023078395 | 0.015525734 | 0.031168831 | 0.010929847 | 521.3920513 | 541.2154008 | 1.49627736 | -0.010544751 |
| 147_AA | ppt_win | -69.57244202 | 12.2362021 | 11.8022093 | 0 | 0.045803089 | 0.080830455 | 0.054407075 | 0.031282593 | 0.067037669 | 143.144884 | 162.9578574 | -4.849886461 | 0.021207215 |