Genetic analysis of European beech populations across precipitation gradients: understanding the adaptive potential to climate change

Dissertation

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy (Ph.D.)
at the Faculty of Forest Sciences and Forest Ecology,
Georg-August University of Göttingen

by

Laura Carolina Cuervo Alarcon
born in Bogota, Colombia

Göttingen, 2017

1. Gutachter: Prof. Dr. Konstantin Krutovsky
2. Gutachter: Prof. Dr. Heiko Becker

Tag der mündlichen Prüfung: 16.03.2017

A mis padres María y Jorge
y mis hermanas Mayis y Ara
por cobijarme con su amor y soporte en esta aventura lejos de casa

A Jhon, porque contigo me atrevo a ser lo mejor de mi misma

A mi frijolito, porque me trajiste de vuelta a lo que realmente importa

To my parents María and Jorge
and my sisters Mayis and Ara,
for embracing me with their love and support in this adventure far away from home

To Jhon, because with you I dare to be the best of myself

To my little bean, because you brought me back to what really matters

For many peoples of black Africa, ancestors are the spirits that live in the tree beside your house or in the cow grazing in the field. The great-grandfather of your great-great-grandfather is now that stream snaking down the mountainside. Your ancestor could also be any spirit that decides to accompany you on your voyage through the world […].
And the ancestral spirits, the ones that help you make your way, are the many grandparents that each of you has. As many as you wish.

-Eduardo Galeano, Mirrors: Stories of Almost Everyone

Acknowledgments

I want to express my deep gratitude to Prof. Dr. Konstantin Krutovsky for accepting me as his PhD student and for his valuable guidance, support and encouragement through my doctoral studies.

I want to thank Prof. Dr. Reiner Finkeldey for accepting me in the first place as a member of the Forest Genetics and Forest Tree Breeding group and helping with establishing collaboration with the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), sharing his research ideas and developing the experimental plan.

I thank Prof. Dr. Heiko Becker for being co-referee of this thesis and Prof. Dr. Andrea Polle for being a member of my thesis committee; their valuable comments and discussions during our meetings helped me improve scientifically. I also thank Prof. Dr. Niko Balkenhol for being member of my examination committee.

I also want to extend my thanks to my counterpart Dr. Matthias Arend from the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL) for providing me the phenotypic data used in the phenotypic association analysis, and also for his support during the field work sampling the adult trees and his helpful comments on a draft version of this thesis.

My warmest gratitude goes to Dr. Markus Müller, for answering all my questions about beech, his scientific advice and helpful comments on a draft version of this thesis.

I also want to thank Dr. Hadrien Lalagüe for kindly providing me the SNP positions and haplotype sequences of beech populations from France, making possible for me to select the haplotype Tag SNPs used in this study.

I am grateful to Prof. Dr. Martin Ziehe and Dr. Elizabeth Gillet for useful discussions on statistical methods and data analysis. I am thankful to Alexandra Dolynska for her valuable technical assistance and teachings on lab work. I also thank Regina Berkeley and Kerstin Quentin for their administrative support.

I thank my office mates Devrim Semizer Cuming and Sinan Hagenah not only for the company, but also for the nice conversations and laughs. My special thanks go to Dr. Natalie Breidenbach, for being my official translator, for her friendship and countless laughs. I thank also Dr. Oleksandra Kuchma and Alexandra Dolynska for making lunch a nice time.

Special thanks to all my former and current colleagues in the Department of Forest genetics and Forest Tree Breeding for the helpful discussions, questions, answers and nice time spent together: Prof. Dr. Hans Heinrich Hattemer, Dr. Ludger Leinemann, Dr. Barbara Vornam, Dr. Fitri Y. Amandita, Dr. Nurlan Torokeldiev, Johanna Ropertz, Melanie Schmitt, Christine Radler, Gerold Dinkel and Marco Winkler.

I acknowledge COLFUTURO and the Administrative Department of Science, Technology and Innovation - COLCIENCIAS for supporting me during my doctoral studies.

I am totally grateful to my great love Jhon Rivera, for being by my side and holding my hand. Last but not least, my deepest gratitude to my parents Maria and Jorge and my sisters Mayis and Ara, for believing in me and encouraging me to pursue my dreams.

Table of content

1. General Introduction **1**

Global Climate Change **1**

Climate change in Europe 2

Climate change in Switzerland 3

Effects of climate change on biodiversity 3

Biology and distribution of European beech **5**

Impact of climate change on European beech 6

Genetic variation and differentiation 8

Neutral and adaptive genetic variation **9**

Neutral Genetic variation 9

Adaptive genetic variation 10

Approaches to detect adaptive variation 10

Objectives **15**

2. Genetic diversity and population structure **16**

Introduction **16**

Materials and methods **18**

Plant material 18

DNA isolation 18

SSR amplification and genotyping 19

Candidate genes and SNPs 20

Data analysis 22

Results **24**

SSRs 24

SNPs 35

Discussion **48**

Genetic variability and linkage disequilibrium (LD) 48

Genetic differentiation and population structure 49

Outlier analysis 50

3. Phenotypic Association analysis **53**

Introduction **53**

Materials and Methods **54**

Plant material 54

Experimental design 54

Phenotypic data ... 55
Statistical analysis of phenotypic data ... 56
SNPs and candidate genes ... 56
Association analysis ... 58
Results ... 59
Phenotypic data ... 59
Phenotypic Association analysis ... 65
Discussion ... 69
Day of Bud burst ... 70
Chlorophyll fluorescence ... 70
Stem Growth ... 71
Association analysis ... 72
Appendix 3-1 ... 75

4. Association of genetic variation with environment ... 76
Introduction ... 76
Materials and methods ... 77
Plant material ... 77
DNA isolation ... 78
Candidate genes and SNPs ... 78
Environmental data ... 78
Environmental association analysis ... 82
Results ... 82
Environmental data ... 82
Environmental association analysis ... 85
Discussion ... 88
Appendix 4-1 ... 93
Appendix 4-2 ... 94

5. Spatial genetic structure, relatedness and parental assignment ... 96
Introduction ... 96
Materials and methods ... 97
Plant material ... 97
DNA isolation and microsatellite loci genotyping ... 97
Relatedness and spatial genetic structure analysis ... 98
Parentage assignment ... 99

Results **99**
Relatedness and spatial genetic structure 99
Parentage assignment 102
Discussion **103**
Appendix 5-1 ..... **106**

6. Synthesis **118**
General results and discussion **118**
Conclusions and perspectives **122**

Summary **123**

Zusammenfassung **125**

Bibliography **127**

List of abbreviations

AMOVA	Analysis of Molecular Variance
EAA	Environmental Association Analysis
EST	Expressed Sequence Tag
FDR	False Discovery Rate
GLM	General Linear Model
GWAS	Genome-wide association study
LD	Linkage Disequilibrium
LFMM	Latent Factor Mixed Model
MAF	Minimum Allele Frequency
MCMC	Markov Chain Monte Carlo
MLM	Mixed Linear Model
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PET	Potential Evapotranspiration
SG	Stem Growth
SNP	Single Nucleotide Polymorphism
SSR	Short Sequence Repeat

List of tables

Table 2-1 Environmental characteristics of the selected populations 18

Table 2-2 Candidate genes and characteristics of the selected SNPs 20

Table 2-3 Diversity parameters for SSR loci 25

Table 2-4 Diversity parameters for sapling and adult populations based on SSR loci 26

Table 2-5 AMOVA based on SSR loci for saplings and adults 27

Table 2-6 List of the SSR loci detected as outliers 34

Table 2-7 Diversity parameters for SNPs 36

Table 2-8 Diversity parameters for saplings and adult populations based on SNPs...............38

Table 2-9 AMOVA based on SNPs for saplings and adults 40

Table 2-10 List of the SNPs detected as outliers 46

Table 3-1 Environmental characteristics of the selected populations 54

Table 3-2 Summary of the number of saplings under different conditions 55

Table 3-3 Candidate genes and characteristics of the selected SNPs 57

Table 3-4 Results of the phenotypic association analysis 66

Table 4-1 Candidate genes and characteristics of the selected SNPs 79

Table 4-2 Abbreviation and description of the geographical and environmental variables....81

Table 4-3 Eigenvalue and variance explained for climatic principal components 83

Table 4-4 Correlation coefficients between the environmental variables and PCs.................84

Table 4-5 Population values for the first three climatic principal components 85

Table 4-6 List of SNPs that significantly correlated with climatic PCs.................................86

Table 5-1 Mean values per population of the relatedness coefficient r_{QG} for pairs of saplings 100

Table 5-2 Percentage of saplings with a tentative parent assigned 102

Table 6-1 List of SNPs that show signatures of selection..120

List of figures

Fig. 1-1 Distribution map of F. sylvatica 5

Fig. 2-1 Linkage disequilibrium between pairs of SSR loci in the saplings 26

Fig. 2-2 Pairwise genetic differentiation with SSRs....................27

Fig. 2-3 Structure analysis of sapling populations based on the SSRs....................28

Fig. 2-4 Structure analysis of adult populations based on the SSRs 29

Fig. 2-5 Neighbor Joining dendrograms based on SSRs....................30

Fig. 2-6 Detection of outlier SSR loci in saplings using Arlequin 31

Fig. 2-7 Detection of outlier SSR loci in saplings using BayeScan 31

Fig. 2-8 Detection of outlier SSR loci in adults using LOSITAN....................32

Fig. 2-9 Detection of outlier SSR loci in adults using Arlequin....................33

Fig. 2-10 Detection of outlier SSR loci in adults using BayeScan....................33

Fig. 2-11 Allele and heterozygosity frequency distribution in adults for SSR likely under selection 35

Fig. 2-12 Linkage disequilibrium between pairs of SNPs in saplings....................39

Fig. 2-13 Linkage disequilibrium between pairs of SNPs in adults 39

Fig. 2-14 Pairwise genetic differentiation estimated with SNPs 40

Fig. 2-15 Population structure of populations of saplings based on SNP data....................41

Fig. 2-16 Population structure of populations of adults based on SNP data 42

Fig. 2-17 Neighbor joining dendrograms based on SNP data 43

Fig. 2-18 Detection of outlier SNPs in saplings using LOSITAN 44

Fig. 2-19 Detection of outlier SNPs in saplings using Arlequin 44

Fig. 2-20 Detection of outlier SNPs in adults using LOSITAN 45

Fig. 2-21 Detection of outlier SNPs in adults using Arlequin....................45

Fig. 2-22 Allele frequency distribution for the outlier SNPs....................47

Fig. 3-1 Day of bud burst of the single populations 60

Fig. 3-2 Day of bud burst over all populations....................61

Fig. 3-3 Physiological responses of the single populations to control and drought treatments measured by chlorophyll fluorescence parameters 61

Fig. 3-4 Responses of mesic and xeric populations to the drought treatment measured by the parameter PI_{tot} 62

Fig. 3-5 Responses over all populations to control and drought treatments measured by PI_{abs} and PI_{tot} 62

Fig. 3-6 Morphological responses of the single populations to treatment and type of soil measured as stem growth in 2013 and 2014 63

Fig. 3-7 Morphological responses of the single populations to treatment and type of soil measured as overall stem growth 2013 - 2014 64

Fig. 3-8 Effect of treatment and type of soil on stem growth over all populations 64

Fig. 3-9 Differences in the responses of mesic and xeric populations to the drought treatment measured by stem growth 65

Fig. 3-10 SNPs showing association in the analysis using all saplings and their effects on normalized chlorophyll fluorescence parameters 67

Fig. 3-11 SNPs showing association in saplings under drought/acidic soil conditions and their effects on the phenotypic traits 67

Fig. 3-12 SNP showing significant association in saplings under control/acidic soil conditions and its effect on PIabs 68

Fig. 3-13 SNPs showing significant association in saplings under drought/calcareous soil conditions and their effects on stem growth 69

Fig. 4-1 Spearman's rank correlation coefficients matrix between environmental variables .83

Fig. 4-2 Relationship between climatic variation at the PC1 and minor allele frequencies for SNPs that were also identified as outliers 87

Fig. 4-3 Relationship between climatic variation at the PC2 and minor allele frequencies for SNPs that were also identified as outliers 88

Fig. 4-4 Relationship between climatic variation at the PC3 and minor allele frequencies for SNPs that were also identified as outliers 88

Fig. 5-1 Distribution of pairwise r_{QG} coefficients in saplings 100

Fig. 5-2 Correlograms showing the correlation coefficient *r* as a function of geographic distance 101

1. General Introduction

Global Climate Change

Human activities are changing the planet and one of the consequences is climate change, which can be defined as changes in climatic global conditions that involve changes in temperature, precipitation and frequency of extreme events such as floods and droughts (Franks & Hoffmann 2012).

During the last century, economic and population growth were the most important causes of an increment in carbon dioxide (CO_2) emissions, mainly due to fossil fuel combustion (IPCC 2014). Furthermore, the concentration of other green house gases such as nitrous oxide (N_2O), methane (CH_4), hydrofluorocarbons and perfluorocarbons, has increased considerably in the atmosphere, especially during the last century (Hartmann *et al.* 2013). It is considered that these green house gases are the main cause of the observed warming in the 20th century, leaving little doubt that human activities are altering the climate (IPCC 2014).

Since the late 19th century, global mean temperature has been increasing, being the first decade of the 21st century the warmest. This trend towards warming has been observed in both the atmosphere and the ocean, and an increment in the temperature of 0.85ºC over the period between 1880-2012 has been estimated (Hartmann *et al.* 2013). Related to this, it has been observed on a global scale that the number of warm days has increased while the number of cold days has decreased (IPCC 2014). Also, rising temperatures have caused a reduction in the amount of ice cover and spring snow cover in the Northern hemisphere, as well as a reduction in the permafrost (Dore 2005; IPCC 2014). Furthermore, an increment in sea level has been observed (IPCC 2014).

The hydrological cycle has also been affected by increasing temperatures, leading to changes in precipitation. First of all, the probability of precipitation falling as rain rather than snow increases, something that has been observed principally in spring and autumn in the Northern hemisphere. As a result, snow pack area is reduced, and since it constitutes an important source of freshwater in spring and summer as snow melts, less water is available during those seasons (Trenberth 2011). Second, changes in the distribution of precipitation were also observed during the 20th century, and even though they exhibit high spatial and temporal variability, some tendencies were observed: precipitation increased in North America, Eurasia and Argentina, while a reduction was observed in southern Europe, the Mediterranean, Africa and the tropics (Dore 2005; Trenberth 2011).

However, the most notorious effect of a warmer climate is the occurrence of more extreme events that increment the risk of floods and droughts. Increasing temperatures result in more water vapor, and thus, more moisture accumulated in the atmosphere, favoring the occurrence of heavy rain. During the last century, more heavy precipitation events and consequently flood incidents were observed (Trenberth 2011), and in some regions of the world, the increment in precipitation was caused by more heavy precipitation (Dore 2005). On the other hand, warming temperatures also increased drought occurrence and severity, since drying of land surface is enhanced as moisture evaporation is accelerated. During the last century, more droughts were caused by increased evapotranspiration due to higher temperatures and changes in precipitation distribution, being Africa, the tropics and subtropics the most affected regions (Trenberth 2011). In general, the global trend indicates that wet areas are becoming wetter and dry areas are becoming drier (Dore 2005).

The rising temperatures, changing patterns in precipitation and extreme events that occurred during the last century are the signatures of global climate change. The emission of green house gases is the principal responsible for the warming of the planet, and different emission scenarios were evaluated by the Intergovernmental Panel on Climate Change (IPCC) for the 21st century: one stringent mitigation scenario (RCP2.6), two intermediate scenarios (RCP4.5 and RCP6.0) and one scenario with very high emissions of green house gases (RCP8.5). Even though the magnitude of the projected climate change depends on the scenario considered, under all of them surface temperature is projected to rise over the 21st century, and this increment shows a strong relationship with emissions of CO_2. This will leave ecosystems and human systems more vulnerable to heat waves, droughts, floods, cyclones, wildfires, ocean acidification and rising of sea level (IPCC 2014)

Climate change in Europe

Temperature in Europe has been increasing, showing an increment of 1.3ºC in the first decade of the 21st century compared to the last half of the 19th century. However, the observed warming has been different over the continent both spatially and temporally: Northern Europe is getting warmer particularly in winter, whereas Southern Europe is getting warmer mainly in summer, and climate projections under all emission scenarios predict that this trend will continue. Likewise, since 1950, hot days, tropical nights and heat waves have

increased in frequency, whereas cold spells and frost days have been reduced. This tendency will continue during this century (Kovats *et al.* 2014)

During the 20th century, annual precipitation showed an increment in the north and a decrease in the south of Europe (Dore 2005), and it is projected that this trend will continue during this century (Kovats *et al.* 2014). Also, a result of warming, precipitation in winter will be more likely rain rather than snow, especially in mountainous areas (Kovats *et al.* 2014). Furthermore, extremes events will be intensified: flood events resulting from heavy precipitation will be more likely to occur in the north and northeast of the continent; and droughts will be more likely to occur in Central and Southern Europe and the Mediterranean (Lehner *et al.* 2006; Kovats *et al.* 2014). Even in regions where it is expected an increase of summer precipitation, soil moisture can be lost due to increased evapotranspiration as a result of warmer temperatures, and this may lead to more severe hydrological droughts (Kovats *et al.* 2014).

Climate change in Switzerland

In the last century, an increment in temperature of 1.2 °C has been observed in Switzerland, and also warmer and drier summers have become more common since the 1970s (Beniston & Goyette 2007). By the end of the 21st century, it is predicted not only an increment of 4°C in minimum and maximum temperatures, but also an increment in the frequency of intense and longer lasting summer warm periods and heat waves (Beniston & Goyette 2007; CH2011 2011). Additionally, since 1990s it has been observed that the persistence of cold events such as cold winter days and nights has been decreasing, a tendency that is expected to continue through the 21st century (Beniston & Goyette 2007; CH2011 2011). Correspondingly, a shift in precipitation from snow to rain is expected during winter, while in summer the amount of rain is projected to decrease, affecting mainly the Alpine region and making dry conditions more likely to occur (CH2011 2011).

Effects of climate change on biodiversity

Climate change is affecting biodiversity in complex ways. Changes in the distribution, abundance, phenology, and migration patterns of different species have been observed. Also, interactions among species have been affected, e.g. plant pests and diseases, disease vectors

and hosts (IPCC 2014; Kovats *et al.* 2014). Likewise, some negative effects have been observed on crop health and productivity, due to altered disease epidemiology and host resistance (Chakraborty *et al.* 2000). For the 21st century, it is projected that crops such as wheat, rice and maize will be affected because of climate change, having an impact on food security (IPCC 2014).

Forest trees are being affected by climate change in several ways. First, rising temperatures are causing changes in phenology. The beginning of growth in spring occurs when a determined chilling sum is met in winter, followed by a determined heat sum in spring; therefore, climate change may delay the satisfaction of chilling requirements in winter or accelerate the satisfaction of heat sum requirements in spring (Aitken *et al.* 2008). Indeed, it has been observed that phenological traits such as flowering are beginning earlier, increasing the risk of frost floral damage and generating mismatches between plants and their pollinators (Schröder *et al.* 2006; Anderson *et al.* 2011; DeLucia *et al.* 2012), which could affect the reproductive synchronicity among populations and long distance gene flow via pollen (Aitken *et al.* 2008).

Second, changes in precipitation patterns that increase the likelihood of extreme events such as floods and droughts, will likely affect forests survival. Even in regions of Europe where is projected an increment in summer precipitation, soil moisture can be lost due to increasing evapotranspiration rates as a result of warmer temperatures, leading to more severe hydrological droughts (Kovats *et al.* 2014). During the growing period of forest trees, water supply has an important influence on the vitality, growth and organic matter production of the forest. Water deficiency during warm months combined with high evapotranspiration can restrict forest growth and survival (Führer *et al.* 2011). For example, in the Iberian Peninsula an increment in defoliation and tree mortality in the last two decades as a consequence of drought has been observed (Carnicer *et al.* 2011).

Third, the interaction between forest trees and other species is being altered. Besides the mismatches observed in the interaction between plants and their pollinators (Schröder *et al.* 2006; Anderson *et al.* 2011; DeLucia *et al.* 2012), interactions between forest trees and ecto-micorrizal fungi and insects is being affected, causing changes their diversity and abundance (Swaty *et al.* 2004; Trotter *et al.* 2008; Stone *et al.* 2010). Thus, climate change will very likely affect forest structure, composition, distribution and productivity, which in turn will affect other species that depend on forests to survive (Allen *et al.* 2010; Zhao & Running 2010; Crookston *et al.* 2010; Chmura *et al.* 2011).

Biology and distribution of European beech

European beech (*Fagus sylvatica* L.) is one of the most important forest trees in Europe. It is a deciduous and monoecious tree that usually reaches 30-40 m tall and has a typical life span of 150-300 years (Houston Durrant *et al.* 2016), reaching sexual maturity at approximately 40 years of age (Packham *et al.* 2012) and maintaining a high growth rate until late maturity (von Wuehlisch 2008). Pollen dispersion occurs by wind, while seed dispersion occurs primarily by gravity and secondary by animals such as rodents and birds (Jensen 1985; Nilsson 1985; Perea *et al.* 2011; Packham *et al.* 2012). At sites with favorable environmental conditions, European beech is the dominant species due to its high competitive ability given by its efficient use of light and shade tolerance, forming a dense canopy under which beech seedlings are more likely to outcompete seedlings of other species (Jahn 1991; von Wuehlisch 2008)

F. sylvatica is broadly distributed over Europe (Fig. 1-1), extending from southern Scandinavia in the north to Sicily in the south, and from Spain in the west to northwest Turkey and eastern Poland in the east (Packham *et al.* 2012; Houston Durrant *et al.* 2016). Its distribution is more concentrated in Central Europe, where more favorable environmental conditions exist (Bolte *et al.* 2007) In Switzerland, *F. sylvatica* is the second most important tree species, covering a wide range of the forested area and being predominant in the sub-montane and lower montane range (Weber *et al.* 2010).

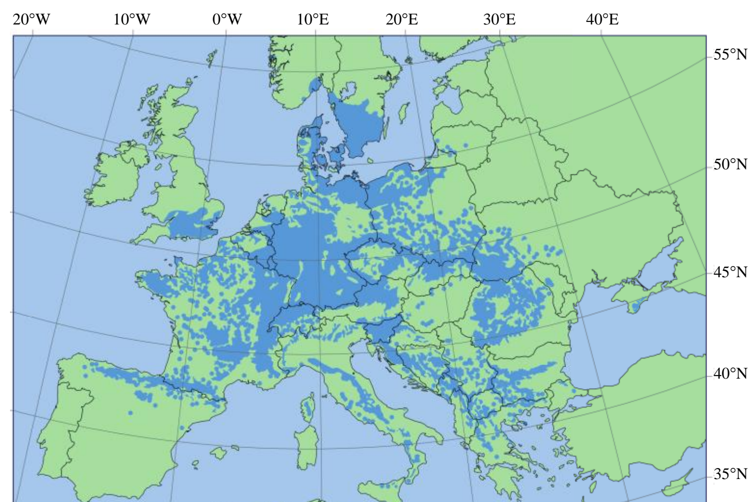


Fig. 1-1 Distribution map of F. sylvatica (blue shaded) in Europe. EUFORGEN 2009, www.euforgen.org

As for other species of beech, the distribution of *F. sylvatica* depends mainly on temperature and moisture availability (Fang & Lechowicz 2006), avoiding extreme

conditions of temperature and precipitation (Jahn 1991). The average annual temperature at the southern and northern limit of beech distribution is 13.5 ºC and 6.6 ºC, respectively; whereas annual precipitation has an average of 906 and 1272 mm at the southern and northern limit of distribution, respectively (Fang & Lechowicz 2006). Low temperatures at the northern and eastern limits of distribution, as well as in high altitudes, are the limiting factor for the growth of beech, because it is susceptible to extreme winter conditions and to spring and autumnal frosts (Jahn 1991; Packham *et al.* 2012). On the other hand, at the southern limit of distribution and in low altitudes water deficit is the limiting factor (Jahn 1991). The combination of low precipitation and high temperatures creating dry conditions can limit the growth of beech, unless the low precipitation is compensated for by high soil moisture or frequent fogs (Jahn 1991).

European beech has low soil nutrient requirements (Jahn 1991). Its optimal growth is reached in humid soils located on calcareous or volcanic rocks; however, it can grow on many types of soil with a pH between 3.5 and 8.5, avoiding soils with deficit or excess of water and too rocky sites or very dense soils where the roots cannot penetrate easily (Jahn 1991; Houston Durrant *et al.* 2016). In general, the more favorable the climate is for European beech, the less is its soil specificity (Jahn 1991).

Impact of climate change on European beech

As a consequence of warming temperatures, changes in the phenology of this species have been observed, registering an advancement of spring bud burst parallel with the global climatic trend (Badeck *et al.* 2004). This increases the probability of late frost damage, affecting especially the survival of seedlings and saplings (Packham *et al.* 2012). Besides, the performance of *F. sylvatica* is significantly influenced by temperature and moisture availability, growing more vigorously where summer temperatures and water stress are lower (Packham *et al.* 2012). Thus, the predicted increment in frequency and duration of summer droughts under climate change will also likely have an effect on the survival and distribution of this species (Gärtner *et al.* 2008; Kramer *et al.* 2010).

Severe drought periods may be harmful for European beech. The root system of this species tends to be shallow, making it susceptible to drought when compared to coniferous stands (Packham *et al.* 2012). Compared to other European forest tree species, it has been found that *F. sylvatica* is very sensitive to drought (Köcher *et al.* 2009), reducing

significantly its growth with increasing drought stress conditions (Scharnweber *et al.* 2011). Furthermore, important physiological functions such as leaf conductance, photosynthetic activity, stem hydraulic conductivity, fine root vitality and nutrient uptake are sensitive to drought in beech (Leuschner *et al.* 2001; Geßler *et al.* 2007; Milad *et al.* 2011). Thus, under a climate change scenario with more frequent droughts, *F. sylvatica* could be overcompeted by more drought-tolerant trees such as *Quercus petraea* and *Pinus sylvestris* (Geßler *et al.* 2007; Friedrichs *et al.* 2009), leading to a reduction in its abundance and changes in its distribution. Simulation studies project that *F. sylvatica* could lose nearly 29% of its habitat under climate change, with a population reduction in the south and expansion in the north, and a shift in distribution towards higher elevations (Kramer *et al.* 2010; CH2014-Impacts 2014).

Despite its susceptibility to drought, it has been suggested that populations of European beech from dry sites could be more drought-tolerant than populations from wet sites. For example, Peuke *et al.* (2002) studied the response to drought treatment of seedlings from populations with different amount of precipitation. They found that the water potential and transpiration rates of seedlings from dry habitats were less affected by drought; furthermore, a low concentration of osmoprotectans such as proline and of the hormone ABA was found in these seedlings (Peuke *et al.* 2002). Additionally, studies on marginal populations from the southern and north-eastern limits of distribution, considered to represent dry conditions, have shown that seedlings from those sites have higher root/shoot ratio under drought conditions, which may facilitate access to soil water (Rose *et al.* 2009); also, their growth is less affected by drought (Thiel *et al.* 2014). Similar results have also been found on adult trees. In the typically xeric Mediterranean environment occurring in Greece, beech trees did not show signs of drought stress in physiological parameters such as leaf water potential and carbon isotopic composition during a three year period including the year 2003, one of the driest and hottest years registered for Europe that affected beech in central Europe (Fotelli *et al.* 2009). Furthermore, dendroecological data also indicate that populations from the dry distribution limit are better able to cope with dry conditions, since they exhibit higher tree-ring growth and are less sensitive to drought than populations at mesic sites (Weber *et al.* 2013).

Genetic variation and differentiation

Genetic studies on beech using different genetic markers such as isoenzymes, RAPDs, AFLPs, microsatellites and SNPs, have found that this forest tree species is characterized by high genetic variability (Sander *et al.* 2000; Emiliani *et al.* 2004; Jump & Peñuelas 2007; Kraj & Sztorc 2009; Pluess & Weber 2012; Müller *et al.* 2015a). As for other forest trees, this high genetic variability is explained by a combination of ecological and life history traits, such as long life, outcrossing breeding system, wide pollen dispersal and large geographic range of distribution (Hamrick *et al.* 1992).

Low to moderate genetic differentiation among populations is another characteristic of *F. sylvatica.* This has been found in different studies across Germany (Sander *et al.* 2000; Rajendra *et al.* 2014; Müller *et al.* 2015a), Italy (Paffetti *et al.* 2012), France (Csilléry *et al.* 2014) and other parts of Europe (Buiteveld *et al.* 2007). Since *F. sylvatica* is mainly an outcrossing tree species, this low differentiation can be explained by high gene flow among populations through pollen dispersal. Indeed, pollen immigration rate has been estimated to be about 75% (Oddou-Muratorio *et al.* 2011; Piotti *et al.* 2012), and pollen dispersal can cover thousands of kilometers, from Germany and North Italy to Catalonia in Spain (Belmonte *et al.* 2008). Even though seed dispersal can cover shorter distances than pollen dispersal, gene flow through seeds can also contribute to low genetic differentiation. A immigration rate of about 20% has been found for seeds (Oddou-Muratorio *et al.* 2011), and seed dispersal of 1400 and 3000 m has been reported (Kunstler *et al.* 2007). Although primarily seed dispersal occurs by gravity, accounting for dispersion few meters away from the mother tree (Millerón *et al.* 2013), longer distances can occur by birds such as nuthatches (*Sitta europaea*), great tits (*Parus major*) and jays (*Garrulus glandarius*) (Nilsson 1985; Perea *et al.* 2011).

The investigation of local spatial genetic structure in beech has shown that exists strong family structure up to distances of 20-110 m (Vornam *et al.* 2004; Jump & Peñuelas 2007; Chybicki *et al.* 2009; Piotti *et al.* 2013), meaning that closer individuals are more genetically related. This is attributed to the gravity dispersal nature of beech seeds, which are released under the canopy of the mother tree. Using genetic markers for parentage analysis, it has been possible to determine that within stands the range of seed dispersal is between 40-50 m (Millerón *et al.* 2013; Bontemps *et al.* 2013), with a mean of 11m (Oddou-Muratorio *et al.* 2011).

Since *F. sylvatica* is one of the most important species in Europe, and morphological and physiological data indicates that that beech populations in dry areas cope better with drought, there has been a great interest in the identification of genetic variation underlying adaptation to changing environmental conditions. Using AFLPs, genetic differences have been found between beech populations growing in sites with different water availability (Pluess & Weber 2012). Recently, the development of SNP markers in climate-related candidate genes has been reported (Seifert *et al.* 2012; Lalagüe *et al.* 2014; Müller *et al.* 2015b), and associations between those SNPs and important climate related traits, such as bud burst (Müller *et al.* 2015a), elevation (Csilléry *et al.* 2014), temperature, precipitation and drought, have been detected (Pluess *et al.* 2016). However, much remains to be known about the genetic adaptive variation in *F. sylvatica* and its implications for the adaptation of this important species to climate change.

Neutral and adaptive genetic variation

Neutral Genetic variation

Neutral genetic variation is genetic variation that does not have an effect on fitness, and thus, is selectively neutral, being influenced by mutation, gene flow and genetic drift. Neutral genetic variation is used to study processes like gene flow, migration and dispersal (Holderegger *et al.* 2006), and also for identification of species and management units in conservation (Hedrick 2001). However, it provides little insight into local adaptation and evolutionary potential (Kirk & Freeland 2011). Among the existent molecular markers, microsatellites are the most commonly used for the study of neutral genetic variation (Holderegger *et al.* 2006; Kirk & Freeland 2011)

Microsatellites, also known as simple sequence repeats (SSRs) or short tandem repeats (STRs), are sequences of 1-6 nucleotides repeated in tandem (Haasl & Payseur 2013). SSRs are very polymorphic, presenting multiple alleles mainly as a result of variability in length i.e., variability in the number of repetitions, rather than variation in sequence (Ellegren 2004). This high genetic variability is due to very high mutation rates when compared to point nucleotide mutations in coding regions (Bhargava & Fuentes 2010). SSRs are widespread in both prokaryotes and eukaryotic genomes (Bhargava & Fuentes 2010). They are mainly located in non-coding regions, and thus, assumed to be neutral and referred to as genomic SSRs (Ellegren 2004; Holderegger *et al.* 2006). However, SSRs can also be found in coding

regions, and thus, could be subject to selection (Ellis & Burke 2007). In plants, SSRs are found only in a low percentage of genes (Varshney *et al.* 2005), and they are mainly obtained from expressed sequence tag databases (EST-SSRs).

Adaptive genetic variation

Adaptive genetic variation is the genetic variation that has an effect on fitness and thus, is subject to natural selection. Like neutral genetic variation, adaptive genetic variation is also affected by neutral processes such as mutation, genetic drift and gene flow; however, the effect of selection is stronger and overpasses the effect of neutral processes (Holderegger *et al.* 2006). The study of genetic adaptive variation is important for conservation purposes, because it is directly involved in the response to environmental changes (Hedrick 2001; Hoffmann & Willi 2008).

Single nucleotide polymorphisms, or SNPs, are valuable markers for the study of genetic adaptive variation in plants (Gailing *et al.* 2009). A SNP is a single base pair change that is variable across the population and whose frequency is at least 1% (Foulkes 2009). Even though SNPs are usually biallelic mainly because of low mutation rate, and thus, have low polymorphism, this is compensated for by their high frequency in the genome: in humans, it has been determined that there is a SNP every 1331 bp. In coding regions, SNPs are the most common polymorphism and the most likely responsible for phenotypic variation (Vignal *et al.* 2002). Indeed, several SNPs within a gene are considered in studies using a candidate gene approach for the detection of adaptive genetic variation.

Approaches to detect adaptive variation

A locus can be considered adaptive if it has an effect on a trait with functional importance that is known or suspected to be under selection, or if it shows signatures of historical selection (Barrett & Hoekstra 2011). Different approaches can be used to detect genetic adaptive variation. On the one hand, the bottom-up approaches aim at identifying loci showing signatures of selection in populations from different environments. Loci showing signatures of selection are identified by high levels of genetic differentiation compared to neutral expectations, i.e., outlier loci, or by their association with environmental variation (Barrett & Hoekstra 2011; Rellstab *et al.* 2015). On the other hand, top-down approaches take

traits that are known to be different between environments and aim at detecting the genetic variability underlying those traits. Genome wide association studies (GWAS) and quantitative trait locus (QTL) are top down approaches (Barrett & Hoekstra 2011).

For the detection of adaptive genetic variation, it is recommended to combine several approaches since they complement and support each other. Thus, if a locus is detected by several approaches, then the locus is very likely to be under selection and consequently, the rate of false positives is also reduced (Rellstab *et al.* 2015).

Outlier approaches

Population genetic structure is defined as differences in genetic variation among populations (Hedrick 2005). F_{ST} is the most common index to estimate population differentiation, and can be defined as the probability that two genes within a population share a common ancestor within that population (Beaumont 2005).

Genetic differentiation can arise due to neutral processes as genetic drift, gene flow and mutation (Rellstab *et al.* 2015). However, natural selection is also expected to change allele frequencies among populations, and thus, influence the amount of genetic differentiation. Lewontin & Krakauer (1973) were the first in propose a test to distinguish loci showing signatures of selection based on $F_{ST.}$ They reasoned that the expected amount of differentiation at different loci should be the same because of the shared demographic history experienced by those loci. Thus, loci in which different alleles are selectively favored in different populations should exhibit larger allele frequency differences than do loci with purely neutral alleles. In contrast, loci that are subject to balancing selection should have a lower level of genetic differentiation than neutral loci (Beaumont 2005). Loci showing significantly higher or lower genetic differentiation than expected under neutrality are called outlier loci, and are candidates to be under selection (Antao *et al.* 2008).

Nowadays, there are several approaches based on Lewontin and Krakauer's test; all of them are based on the idea that demographic factors affect the genome in a similar fashion while selection effects are locus-specific. However, one of the drawbacks of outlier approaches is that they can produce false positives, since demographic effects can be confounded with selection (Schoville *et al.* 2012; Vitti *et al.* 2013). To overcome this problem, outlier tests incorporate specific assumptions about demographic history (Antao *et al.* 2008; Foll & Gaggiotti 2008; Excoffier *et al.* 2009). Additionally, to avoid skews in the

estimation of F_{ST} that could lead to false positives or false negatives, it is convenient to use a large number of markers (Nielsen 2005; Beaumont 2005).

Other drawback of outlier tests is that they have little power to detect subtle differences in allele frequencies, as occurs in the presence of high gene flow counteracting selection or in the case of polygenic additive effects (Rellstab *et al.* 2015). Besides, even though outlier tests identify loci that could be under selection or linked to loci under selection, they do not provide insight into the environmental factors that cause selection (Schoville *et al.* 2012; Rellstab *et al.* 2015)

Phenotypic Association Analysis

According to Barrett & Hoekstra (2011), an adaptive allele is an allele that has an effect on a phenotypic trait and produces an increment in fitness. Phenotypic association studies test for associations between phenotypic traits and allelic variation at many loci (Anderson *et al.* 2011); they consider the genotype as the predictor variable and the phenotypic trait as the dependent variable (Foulkes 2009).

Depending on the approach used, association analysis can be classified into two types: on the one hand, genome wide association studies (GWAS) seek to identify causal variants throughout the genome, analyzing hundreds of thousands of SNPs. On the other hand, candidate gene approaches use SNPs within a candidate gene whose function is known or suspected to affect the trait of interest (Balding 2006). One of the advantages of a candidate gene approach is that they provide a direct link to particular candidate genes, and thus, they are less demanding in terms of the number of markers required (Ingvarsson & Street 2011; Franks & Hoffmann 2012). However, this also makes it a limited approach, because it is restricted to genes thought to be involved in the trait of interest and may ignore other genes that could be of relevance or nonidentified candidate genes (Ingvarsson & Street 2011).

The power of phenotypic association analysis to detect an association between a SNP and a trait depends on the phenotypic variance explained by the SNP. The phenotypic variance is determined by how strongly the two alleles differ in their phenotypic effect i.e., their effect size. Some traits are controlled by a small number of loci with large effect sizes. Other traits are controlled by many rare alleles, each having a large effect on the phenotype, or by many common alleles with a small phenotypic effect (Korte & Farlow 2013). One of the limitations of association analysis is their lack of power to detect loci with small effect

size or with low allele frequency; to overcome this problem, large sample sizes are recommended. In humans, most analyses require several thousands of individuals to detect associations, since a large number of small effect loci are found (Pearson & Manolio 2008; Korte & Farlow 2013).

Since association approaches are population based, they need to control for population structure: false positives may occur when phenotypic variation for the trait of interest overlaps with patterns of population structure (Anderson *et al.* 2011; Brachi *et al.* 2011). Therefore, even loci that are unrelated to the trait will show association because of the confounding effects of population structure (Ingvarsson & Street 2011). Relatedness among individuals is another confounding factor that could lead to false associations. This is because related individuals share alleles, causing a higher correlation of the phenotypic trait among individuals from the same family (Foulkes 2009). Currently, there are mixed models that account for both population structure and genetic relatedness to avoid false positives in association analysis (Korte & Farlow 2013).

Environmental Association Analysis

The basic assumption of environmental association analysis (EAA) is that natural selection along an environmental gradient generates changes in allele frequencies (Schoville *et al.* 2012). Only loci under selection show changes in allele frequency, whereas neutral loci do not show any change because they are not affected by natural selection (Holderegger *et al.* 2010). Thus, the goal of EAA is to identify associations between allele frequencies and environmental variables (Rellstab *et al.* 2015). An advantage of EAA is that they are more sensitive to detect subtle changes in allele frequencies caused by weak selection, as in the case of polygenic traits or under high gene flow (Stephan 2016). Besides, EAA incorporate directly the environmental variables assumed to be responsible for selection (Schoville *et al.* 2012). However, a limitation is the coarse spatial resolution of current climate data sources and their integration only over a certain period of time, ignoring small scale heterogeneity and leading to spatial and temporal interpolations (Rellstab *et al.* 2015)

Like outlier and phenotypic association approaches, false positives may occur in EAA when there is population structure or isolation by distance. This is because covariation of geographic distances and environmental gradients usually occurs, and due to restricted gene flow and genetic drift, allele frequencies at neutral loci will randomly change with the

distance and thus, will change indirectly with the environmental gradient (Holderegger *et al.* 2010). Therefore, it is important to correct for neutral genetic structure or spatial autocorrelation, because correcting for both of them could be very conservative and lead to false negatives (Schoville *et al.* 2012; Rellstab *et al.* 2015). Among the existing methods to detect EAA, the mixed effects models are powerful because they control for the effects of neutral genetic structure treating allele frequencies as response variables, environmental factors as fixed factors and neutral genetic structure as a random factor (Rellstab *et al.* 2015)

Objectives

The present study aims to investigate the genetic basis of adaptation of European beech to different environmental conditions using a candidate gene approach. For this purpose, beech populations along precipitation gradients in Switzerland were selected. Samples of adult trees and saplings were collected, and the saplings were additionally subjected to a controlled drought stress experiment.

The main objectives of the present study are:

- to assess genetic diversity and population structure at potentially neutral markers i.e., microsatellites,
- to assess genetic diversity and population structure at potentially adaptive markers i.e., SNPs in climate-related candidate genes,
- to identify genetic markers potentially under selection by conducting outlier analyses,
- to assess the response of saplings to drought conditions by evaluating morphological and physiological traits,
- to detect potentially adaptive genetic markers by conducting phenotypic association analyses between SNPs and morphological and physiological traits assessed in the drought experiment conducted on the saplings,
- to find potentially adaptive genetic markers by conducting environmental association analyses between SNPs and environmental variables.

2. Genetic diversity and population structure

Introduction

Climate change scenarios predict not only an increment in annual temperatures, but also changes in the patterns of precipitation, increasing the risk of extreme events such as floods and droughts (Trenberth 2011). For Central Europe, it has been observed an increment in the temperature of 1.3°C during the first decade of the 21st century compared to the last half of the 19th century; additionally, an increment in the duration and intensity of droughts has also been observed (Kovats *et al.* 2014). These changes in climate will very likely affect the survival of forest trees (Allen *et al.* 2010), altering the composition and distribution of forests (Crookston *et al.* 2010; Chmura *et al.* 2011).

Fagus sylvatica, or European beech, is one of the most important forest tree species in Europe (von Wuehlisch 2008). It is broadly distributed in the continent, covering an area spanning from the North of Sicily in Italy to Southern Norway and Sweden, and from the Cantabrian Mountains in Spain to the Carpathians and Balkan Mountains in Ukraine, Romania and Bulgaria. Its distribution is denser in Central Europe, where moderate conditions in soil moisture and temperature occur (Bolte *et al.* 2007). In Switzerland, beech covers a wide range of the forested area, being dominant in the sub-montane and lower montane range (Weber *et al.* 2010).

However, the extent of the effect of climate change on *F. sylvatica* is still uncertain. Some studies have reported that drought periods may be harmful, affecting nutrient uptake and reducing growth (Geßler *et al.* 2007; Piovesan *et al.* 2008; Scharnweber *et al.* 2011). Furthermore, under a climate change scenario, beech could lose its habitat and be overcompeted by more drought tolerant trees such as *Quercus petraea* and *Pinus sylvestris* (Geßler *et al.* 2007; Friedrichs *et al.* 2009). However, morphological and physiological data indicate that European beech provenances from dry sites could be more drought tolerant than provenances from wet sites (Peuke *et al.* 2002; Dittmar *et al.* 2003; Rose *et al.* 2009; Thiel *et al.* 2014). Additionally, there is genetic evidence suggesting that populations growing in environments with different water availability are under divergent selection (Pluess & Weber 2012).

Genetic studies on beech using isozymes, RAPDs, AFLPs and microsatellites (SSRs) as genetic markers have found that this forest tree species is characterized by high genetic

variability, high gene flow and low population structure (e.g., Sander *et al.* 2000; Emiliani *et al.* 2004; Jump & Peñuelas 2007; Kraj & Sztorc 2009; Pluess & Weber 2012). However, those markers have limited potential for the study of adaptation. In particular, SSR markers are mainly located in non-coding regions (genomic SSRs) and thus, considered to represent neutral genetic variation, i.e., not being under selection (Holderegger *et al.* 2006). However, some SSRs located in coding regions (EST-SSRs) could be under selection (Ellis & Burke 2007). Instead, single nucleotide polymorphisms (SNPs) are the most common polymorphism in genes, and thus, they are considered to be a suitable approach to study adaptive genetic variation because they are directly linked to coding sequences that can be subject to selection (Morin *et al.* 2004). Recently, the development of SNP markers in climate-related candidate genes for *F. sylvatica* has been reported (Seifert *et al.* 2012; Lalagüe *et al.* 2014; Müller *et al.* 2015b), but so far, only few studies have exploited these data to detect genetic adaptive variation on beech (Csilléry *et al.* 2014; Müller *et al.* 2015a; Pluess *et al.* 2016).

The F_{ST} outlier tests are among the most commonly used methods to detect adaptive genetic variation. These tests rely on the assumption that non-selective processes have the same effect on all the loci of the genome, while selection will affect only certain loci (Lewontin & Krakauer 1973). Thus, loci with genetic differentiation (measured by the F_{ST} parameter) higher or lower than expected under neutrality are considered to be under positive or balancing selection, respectively (Vitti *et al.* 2013). However, one of the disadvantages of outlier detection tests is that they can produce false outliers due to population structure and other confounding effects such as migrations, demographic expansions and bottlenecks (Schoville *et al.* 2012; Vitti *et al.* 2013). Different approaches are advised to address this problem (see, for instance, Schoville *et al.* 2012). Signatures of adaptive processes are not always distinguishable from the genomic background. To find thresholds for selectively neutral variation it is recommended to carefully select selectively neutral markers using genome-wide multiple markers and to compare multiple loci. The selectively neutral markers will capture the genome-wide effect of demography on the genetic variation, and loci departing from that pattern will indicate regions under selection (Nielsen 2005; Li *et al.* 2012). A second alternative is to combine different methods, each one with its own demographic assumptions (Li *et al.* 2012). Loci appearing as outliers when considering different demographic scenarios will be more likely to be real candidate loci under selection.

In this study, by using SSR and SNP markers, the patterns of genetic variability and genetic structure among populations of *F. sylvatica* occurring in two precipitation gradients

were addressed. Furthermore, different approaches were used for the detection of outlier loci that could be related to the different environmental conditions in which the populations occur.

Materials and methods

Plant material

Twelve populations of *F. sylvatica* located in the Rhone and Rhine valleys in Switzerland were used in this study (Table 2-1). The populations were located at similar elevations (from 550 to 850 m above sea level), with a mean annual temperature between 8.9 and 9.2 °C. The mean annual precipitation ranged between 849 and 1334 mm in the Rhine valley, and between 603 and 1012 mm in the Rhone valley (Table 2-1). In a first stage, 16-31 adult trees about 50 m apart from each other were selected per population, and 2-4 saplings with a size of ~20 cm underneath them were sampled. In a second stage, 25 adult trees about 50 m apart from each other were sampled per population. In total, leaves from 300 adult trees and from 755 saplings were collected.

Table 2-1 Environmental characteristics of the selected populations

Valley	Population	*N* Adults	*N* Saplings	Position	Elevation, m.a.s.l	Mean annual temperature, °C	Mean annual precipitation, mm
Rhine	Felsberg	25	62	46°51'N, 9°28'E	650-800	10.0	849
	Chur	25	63	46°52'N, 9°32'E	700-800	10.0	849
	Malans	25	64	46°59'N, 9°34'E	600-700	10.1	1114
	Mastrils	25	62	46°58'N, 9°32'E	550-650	10.1	1114
	Sargans	25	63	47°3'N, 9°26'E	650-750	10.1	1334
	Mels	25	60	47°3'N, 9°24'E	650-750	10.1	1334
Rhone	Ardon	25	63	46°13'N, 7°14'E	750-850	10.1	603
	Chamoson	25	64	46°12'N, 7°12'E	750-850	10.1	603
	Saxon	25	64	46°8'N, 7°11'E	700-800	10.1	603
	Martigny	25	64	46°6'N, 7°6'E	500-700	10.1	855
	Collombey	25	63	46°16'N, 6°56'E	550-650	9.8	1012
	Ollon	25	63	46°18'N, 6°59'E	600-700	9.8	1012

N - number of individuals sampled. Climate data were taken from nearby METEO SWISS stations (distance ≤ 10km) for the 1981-2010 period.

DNA isolation

DNA was isolated from dry leaves using the DNeasy™ 96 Plant Kit (Qiagen, Hilden, Germany). The amount and quality of the DNA were examined using electrophoresis in agarose gel at 1% and 1X TAE as running buffer. DNA was stained with Roti®-Safe GelStain

(Roth, Karlsruhe, Germany), visualized by UV illumination, and compared with a Lambda DNA size ladder (Roche, Mannheim, Germany).

SSR amplification and genotyping

Individuals were genotyped at 13 SSR loci. Ten of them are supposedly selectively neutral random genomic SSRs representing noncoding regions. Six of them were originally developed for *F. sylvatica: FS3-04* (Pastorelli *et al.* 2003), *msf11* (Vornam *et al.* 2004), *csolfagus_06*, *csolfagus_19* (Lefèvre *et al.* 2012), *Fagsyl_002929* and *Fagsyl_003994* (Pluess & Määttänen 2013). Four markers - *sfc0018*, *sfc0161*, *sfc1063* and *sfc1143* - were originally developed for *F. crenata* (Asuka *et al.* 2004). The other three SSR loci - *GOT066*, *FIR065* and *FIR004* - are EST-linked (EST-SSRs). They were originally developed for *Quercus robur* (Durand *et al.* 2010), and successfully used for *F. sylvatica* in this study.

The PCR amplifications were performed using fluorescent dye labeled primers as follows: 6-carboxyfluorescein (FAM) dye for *mfs11*, *sfc0161*, *sfc1063*, *csolfagus_06*, *csolfagus_19*, *Fagsyl_003994* and *FIR004*; and 6-hexachlorofluorescein (HEX) dye for *sfc0018*, *sfc1143*, *Fagsyl_002929*, *GOT066*, *FIR065* and *FS3-04*. This allowed us to assemble four different PCR amplification multiplexes. The 1st multiplex was composed of the *FS3-04* and *msf11* markers, the 2nd multiplex - all four *sfc* markers, the 3rd - the *csolfagus* and *Fagsyl* markers, and the 4th - all three EST markers. The PCR amplifications were performed in a total volume of 15 µL containing 2 µL of genomic DNA (about 10 ng), 1X reaction buffer (0.8 M Tris-HCl pH 9.0, 0.2 M $(NH_4)_2SO_4$, 0.2% *w/v* Tween-20; Solis BioDyne, Tartu, Estonia), 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.3 µM of each forward and reverse primer and 1 unit of *Taq* DNA polymerase (HOT FIREPol® DNA Polymerase, Solis BioDyne, Tartu, Estonia). The amplification conditions were as follows: an initial denaturation step at 95 °C for 15 min, followed by 30 cycles consisting of a denaturing step at 94 °C for 1 min, an annealing step at 55 °C (first, second and third multiplexes) or at 47 °C (EST multiplex) for 30 s and an extension step at 72 °C for 1 min. After 30 cycles, a final extension step at 72 °C for 20 min was included. The PCR fragments were separated and sized on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster City, USA). The GS 500 ROX™ (Applied Biosystems, Foster City, USA) was used as an internal size standard. The genotyping was done using the GeneMapper 4.1® software (Applied Biosystems, Foster City, USA).

Candidate genes and SNPs

SNPs in candidate genes involved in phenology and drought stress tolerance from previously published studies for *F. sylvatica* were selected (Seifert *et al.* 2012; Lalagüe *et al.* 2014; Müller *et al.* 2015b). For the candidate genes that contained several SNPs, linkage disequilibrium (LD) blocks were identified by using the software htSNPer 1.0 (Ding *et al.* 2005) and a subset of SNPs representing the majority of haplotypes (haplotype tag SNPs) were selected for further genotyping. In addition, SNPs showing signatures of natural selection in previous studies (Csilléry *et al.* 2014; Müller *et al.* 2015a) were also selected. Finally, 24 genes and 76 SNPs (21 non-synonymous, 27 synonymous and 28 non-coding SNPs) were selected for genotyping (Table 2-2). Nucleotide sequences neighboring selected SNPs were sent to LGC Genomics Ltd. for primer design and SNP genotyping using the PCR-based KASP^{TM} genotyping assay (Hoddesdon, UK).

Table 2-2 Candidate genes and characteristics of the selected SNPs

Gene	SNP name	Type	Reference
Aldehyde dehydrogenase	ALDH_1	Non-coding	Seifert *et al.* 2012
	ALDH_2	Non-Synonymous	
	ALDH_3	Non-Synonymous	
	ALDH_4	Synonymous	
Isocitrate dehydrogenase	IDH_1	Synonymous	
	IDH_3	Non-coding	
	IDH_4	Synonymous	
Ascorbate peroxidase	APX1_1	Synonymous	
	APX1_2	Non-coding	
	APX4_1	Non-coding	
	APX4_2	Non-Synonymous	
Early responsive to dehydration	ERD	Non-coding	
Dehydrin	Dhn_1	Non-Synonymous	
	Dhn_2	Non-Synonymous	
Glutathione peroxidase	GPX	Non-Synonymous	
Phytochrome B	PhyB	Synonymous	
Cysteine proteinase	CysPro_118	Synonymous	Müller et al. 2015
	CysPro_202	Synonymous	
	CysPro_728	Non-coding	
	CysPro_783	Non-coding	
Chloroplast Chaperonin like	CP10_65	Synonymous	
	CP10_67	Non-Synonymous	
	CP10_377	Non-coding	
	CP10_442	Non-coding	
	CP10_503	Synonymous	
	CP10_749	Synonymous	
	CP10_1317	Non-coding	
	CP10_1428	Non-Synonymous	

Gene	SNP name	Type	Reference
Dof zinc finger protein	DAG_81	Non-coding	
	DAG_289	Non-coding	
	DAG_1059	Synonymous	
Histone 3	His3C1_292	Non-coding	
	His3C2_104	Synonymous	
	His3C2_186	Non-coding	
	His3C2_260	Synonymous	
NAC transcription factor	NAC_854	Non-Synonymous	
	NAC_962	Synonymous	
	NAC_1300	Non-coding	
Protein phosphatase 2C	PP2C_315	Non-Synonymous	
	PP2C_391	Synonymous	
	PP2C_791	Non-Synonymous	
	PP2C_941	Non-coding	
	PP2C_1200	Synonymous	
Xyloglucan endotransglucosylase/hydrolase 23	7_258	Non-coding	Lalagüe et al. 2014
	7_520	Non-coding	
Short chain alcohol dehydrogenase	17_880	Non-coding	
	17_1081	Non-coding	
Potassium transporter 2	39_256	Synonymous	
	39_282	Non-Synonymous	
CRT/DRE binding factor	50_39	Non-Synonymous	
	50_232	Synonymous	
	50_320	Non-coding	
s-adenosyl-l-homocysteine hydrolase	52_1_235	Non-Synonymous	
	52_1_249	Non-Synonymous	
	52_1_368	Synonymous	
Glyceraldehyde 3-phosphate dehydrogenase	68_277	Non-Synonymous	
	68_313	Non-coding	
Light-harvesting complex II protein	88_1_450	Non-coding	
	88_1_727	Synonymous	
	88_1_803	Non-Synonymous	
Catalase	91_2_57	Synonymous	
	91_2_141	Synonymous	
	91_2_231	Synonymous	
	91_2_448	Non-coding	
	91_2_479	Non-coding	
	91_2_504	Non-coding	
1-aminocyclopropane-1-carboxylate oxidase	92_166	Non-coding	
	92_352	Non-Synonymous	
	92_630	Non-coding	
Cytosolic class I small heat-shock protein	110_1_111	Non-Synonymous	
	110_1_293	Synonymous	
	110_1_423	Non-Synonymous	
	110_1_450	Non-Synonymous	
Pectin methylesterase	154_2_137	Synonymous	
	154_2_371	Synonymous	
	154_2_617	Synonymous	

Data analysis

Genetic variability and linkage disequilibrium (LD)

Population diversity parameters such as observed heterozygosity (H_o), expected heterozygosity (H_e) and the fixation index (F_{IS}), as well as deviation from Hardy-Weinberg equilibrium, were calculated for both SSRs and SNPs using the GenAlEx 6.5 software (Peakall & Smouse 2006, 2012). Allelic richness was calculated for SSRs accounting for differences in sample size with the HP-Rare program (Kalinowski 2005) using a sample size of 50 individuals. In addition, the MICRO-CHECKER software (Van Oosterhout *et al.* 2004) was used to identify genotyping errors, such as null alleles, in SSR data. Differences in genetic diversity parameters between adults and saplings and between regions were tested for significance using the FSTAT 2.9.3.2 software (Goudet 1995). The GENEPOP 4.2 program (Raymond & Rousset 1995; Rousset 2008) was used to test for LD between pairs of the SSR loci and between pairs of the SNP loci using 10000 dememorizations, 1000 batches and 10000 iterations per batch for Markov chain parameters.

Population genetic structure and differentiation

To assess genetic differentiation, Hedrick's standardized G''_{ST} (Meirmans & Hedrick 2011) based on SSRs and SNPs was calculated for pairs of populations and for all the populations with the GenAlEx 6.5 software (Peakall & Smouse 2006, 2012) using 999 permutations. Since negative values of G''_{ST} can occur when heterozygosity is high, they were interpreted as zero as recommended by Meirmans & Hedrick (2011). Additionally, using the same software, an analysis of molecular variance AMOVA was done with 999 permutations. Population structure was inferred with SSRs and SNPs using the Bayesian approach implemented in the STRUCTURE 2.3.4 software (Pritchard *et al.* 2000). The admixture model with correlated allele frequencies was used. We used 100000 iterations for both the MCMC (Markov chain Monte Carlo) burn-in period and the following MCMC. We tested from 1 to 20 possible populations or clusters (K), using 20 iterations for each of them. The most likely number of clusters was determined considering K with the highest value of mean posterior probability of the data (LnP (D)) as the highest likelihood number of clusters, and also according to the ΔK method proposed by Evanno *et al.* (2005), which is implemented in the STRUCTURE HARVESTER 0.6.94 software (Earl & vonHoldt 2012). The CLUMPAK

software (Kopelman *et al.* 2015) was used for summation and graphical representation of the results obtained by STRUCTURE.

Dendrograms based on the $(\delta\mu)^2$ genetic distance (Goldstein *et al.* 1995) for SSRs, and Nei's standard genetic distance (Nei 1972) for SNPs, were constructed using the neighbor joining clustering method and the Populations 1.2.31 software (Langella, 1999). Bootstrap values across loci were based on 1000 permutations. The dendrograms were visualized using the FigTree 1.4.1 software (Rambaut, 2014).

Outlier analysis

For detection of the outlier SSR and SNP loci that could be under selection in saplings and adults, three different approaches with different demographic assumptions were used, and their results were compared. The first approach was developed by Beaumont & Nichols (1996) and implemented in the LOSITAN software (Antao *et al.* 2008). This approach determines the expected distribution of F_{ST} vs. H_E under an island model of migration assuming neutrality of the loci. The analysis was done using 200,000 simulations, a confidence interval of 95% and a FDR of 0.1. Both the stepwise mutation model and the infinite allele model were used with SSR data, whereas the infinite allele model was used with SNP data. To run LOSITAN we used a procedure typically used in the similar studies (e.g., Krutovsky *et al.* 2009). LOSITAN was run first using all loci to estimate the mean neutral F_{ST}. After the first run, all loci outside the 95% confidence interval were removed, and using only putatively neutral loci that were not removed, LOSITAN was run again to estimate a second mean neutral $F_{ST.}$ Finally, a third run was done using all loci and the second mean neutral F_{ST}. This procedure lowers the bias when estimating the mean neutral F_{ST} by removing, at the end of the first run, the most extreme loci from the estimation (Antao *et al.* 2008). LOSITAN analysis was done taking into account the entire set of populations, and also for each region (Rhine or Rhone) separately.

The second approach is implemented in the Arlequin 3.5 software (Excoffier & Lischer 2010) and is similar to the one implemented in LOSITAN, but considers a hierarchical island model, in which populations exchange more migrants within groups than between groups (Excoffier *et al.* 2009). To use this approach populations of saplings and adults were grouped hierarchically according to the region; furthermore, populations of saplings were also grouped according to the groups revealed by STRUCTURE (see results). Then, 50000

simulations were carried out, using 10 groups of 100 demes as running conditions. Adjustment of P values for a false discovery rate of FDR = 0.1 was done using the Benjamini & Hochberg (1995) method implemented in the R script “p.adjust” (R Core Team 2016).

The third approach is implemented in the BayeScan 2.1 software (Foll & Gaggiotti 2008). It assumes that populations diverged from an ancestral gene pool, and their allele frequencies show different degrees of differentiation from it. To identify outlier loci, BayeScan evaluates the difference in allele frequencies between each subpopulation and the ancestral gene pool by measuring a subpopulation specific F_{ST} coefficient. Each F_{ST} is decomposed into a population-specific (beta) component and a locus-specific component (alpha). If alpha mainly explains the observed differentiation, then departure from neutrality (selection) is assumed. Running conditions used in BayeScan were as follows: a burn-in period with 50000 iterations, a thinning interval of 10, a sample size of 5000 and 20 pilot runs with 5000 iterations each, for a total of 100000 iterations. A locus was considered outlier if its q value was less than FDR < 0.05 or 0.1. BayeScan analysis was done taking into account the entire set of populations, and also for each region separately.

The outlier loci identified by two or more approaches were considered as true outliers under selection. The outlier loci detected by only one of the approaches were considered to be likely false.

Results

SSRs

Genetic diversity and linkage disequilibrium (LD)

Genetic diversity levels were mostly high, but different among SSR loci. In both saplings and adults, loci FS3-04 and GOT066 had the lowest number of alleles, whereas loci Sfc0161 and csolfagus_19 the highest (Table 2-3). Similarly, GOT066 had the lowest observed (H_o) and expected heterozygosity (H_e), while loci csolfagus_06 and csolfagus_19 had the highest, in both saplings and adults. No loci showed evidence of null alleles; the fixation indices (F_{IS}) were close to zero and no significant deviations from Hardy-Weinberg equilibrium were found, except for locus Fagsyl_003994 that presented a significant excess of heterozygous in the saplings, and loci Sfc0018 and FIR004 that also presented a significant excess of heterozygous in adults. In general, EST-SSRs demonstrated lower genetic diversity than genomic SSRs (Table 2-3).

Table 2-3 Diversity parameters for 13 SSR loci genotyped in saplings and adults

Locus	Saplings				Adults			
	N_a	H_o	H_e	F_{IS}	N_a	H_o	H_e	F_{IS}
Genomic SSRs								
FS3-04	3.333	0.509	0.502	-0.014	2.417	0.447	0.471	0.052
msf11	7.667	0.626	0.638	0.019	6.500	0.627	0.631	0.006
Sfc0018	8.917	0.649	0.664	0.020	6.833	0.637	0.658	0.033*
Sfc0161	11.000	0.767	0.786	0.024	9.250	0.763	0.782	0.025
Sfc1063	8.417	0.795	0.798	0.003	7.667	0.787	0.799	0.015
Sfc1143	8.833	0.732	0.736	0.007	7.833	0.763	0.765	0.003
Fagsyl_002929	7.750	0.691	0.677	-0.021	5.833	0.640	0.663	0.035
Fagsyl_003994	8.500	0.703	0.725	0.031*	7.500	0.713	0.723	0.014
csolfagus_06	9.417	0.835	0.832	-0.004	8.417	0.823	0.842	0.023
csolfagus_19	10.417	0.843	0.830	-0.016	9.583	0.823	0.833	0.012
Mean	**8.417**	**0.715**	**0.718**	**0.002**	**7.183**	**0.702**	**0.717**	**0.020**
EST-SSRs								
FIR004	6.250	0.476	0.482	0.013	4.917	0.400	0.463	0.138***
FIR065	4.167	0.692	0.666	-0.039	4.000	0.743	0.680	-0.096
GOT066	2.333	0.107	0.099	-0.078	2.000	0.080	0.076	-0.052
Mean	**4.250**	**0.425**	**0.416**	**-0.035**	**3.639**	**0.408**	**0.406**	**-0.074**
Grand mean	**7.455**	**0.648**	**0.649**	**-0.004**	**6.365**	**0.634**	**0.645**	**0.016**

N_a – mean number of alleles, H_o – observed heterozygosity, H_e – expected heterozygosity, F_{IS} – fixation index, **P*<0.05, ***P*<0.01, ****P*<0.001

Analysis of genetic diversity revealed no significant differences between saplings and adults (A = 6.36 vs. 6.37, P = 0.9; H_e = 0.649 vs. 0.645, P = 0.6; Table 2-4). Likewise, there were no significant differences between the two regions neither in the saplings (A = 6.49 vs. 6.23, P = 0.3; H_e = 0.653 vs. 0645, P = 0.1; Table 2-4) nor the adults (A = 6.59 vs. 6.14, P = 0.1; H_e = 0.646 vs. 0.644, P = 0.8; Table 2-4). Slight differences were observed between populations (Table 2-4). In the saplings, the allelic richness ranged between 5.5 and 6.61, while in the adults it was slightly higher and ranged between 5.77 and 6.92 (Table 2-4). In both adults and saplings, Chamoson and Mels were the populations with the lowest and the highest allelic richness, respectively. In the saplings, the lowest observed and expected heterozygosity were found in Mastrils (H_o = 0.628; H_e = 0.638), and the highest in Mels (H_o = 0.683; H_e = 0.671) (Table 2-4). In the adults, the lowest observed heterozygosity was found in Chur (H_o = 0.591) and the lowest expected heterozygosity in Ollon (H_e = 0.614), although Chur had the second lowest H_e (Table 2-4). Both observed and expected heterozygosity were the highest in Sargans (H_o = 0.658; H_e = 0.672) (Table 2-4). The F_{IS} indices were close to zero, and no significant deviations from Hardy-Weinberg equilibrium were found, except for the adult trees in the Saxon population.

Table 2-4 Diversity parameters for sapling and adult populations based on 13 SSR loci

Population	Saplings				Adults			
	A	H_o	H_e	F_{IS}	A	H_o	H_e	F_{IS}
Rhine								
Felsberg	6.49	0.646	0.652	0.008	6.69	0.646	0.665	0.029
Chur	6.52	0.639	0.645	0.010	6.15	0.591	0.615	0.040
Malans	6.49	0.653	0.657	0.006	6.85	0.646	0.649	0.004
Mastrils	6.28	0.628	0.638	0.016	6.08	0.634	0.623	-0.018
Sargans	6.56	0.667	0.654	-0.019	6.85	0.658	0.672	0.020
Mels	6.61	0.683	0.671	-0.020	6.92	0.646	0.655	0.014
Mean	**6.49**	**0.653**	**0.653**	**0.000**	**6.59**	**0.637**	**0.646**	**0.015**
Rhone								
Ardon	6.45	0.644	0.641	-0.004	6.23	0.637	0.655	0.029
Chamoson	5.50	0.650	0.641	-0.015	5.77	0.637	0.632	-0.008
Saxon	6.23	0.640	0.645	0.006	6.08	0.628	0.657	0.046*
Martigny	6.39	0.651	0.646	-0.007	6.08	0.625	0.641	0.025
Collombey	6.38	0.628	0.644	0.026	6.54	0.658	0.664	0.008
Ollon	6.44	0.646	0.650	0.007	6.15	0.606	0.614	0.012
Mean	**6.23**	**0.643**	**0.645**	**0.002**	**6.14**	**0.632**	**0.644**	**0.013**
Grand mean	**6.36**	**0.648**	**0.649**	**-0.004**	**6.37**	**0.634**	**0.645**	**0.016**

N – sample size, A – allelic richness, H_o – observed heterozygosity, H_e – expected heterozygosity, F_{IS} – fixation index, *P<0.05

In the saplings, significant LD was observed for 19,2% (15 pairs) of all the possible pairs of SSR loci (Fig. 2-1). In contrast, for populations of adults, only the Sfc0018-FIR065 pair (0.013%) showed significant LD. This pair demonstrated LD also in the saplings. There are no linkage mapping data for the studied loci, therefore, it is impossible to see if observed LD is due to the close linkage.

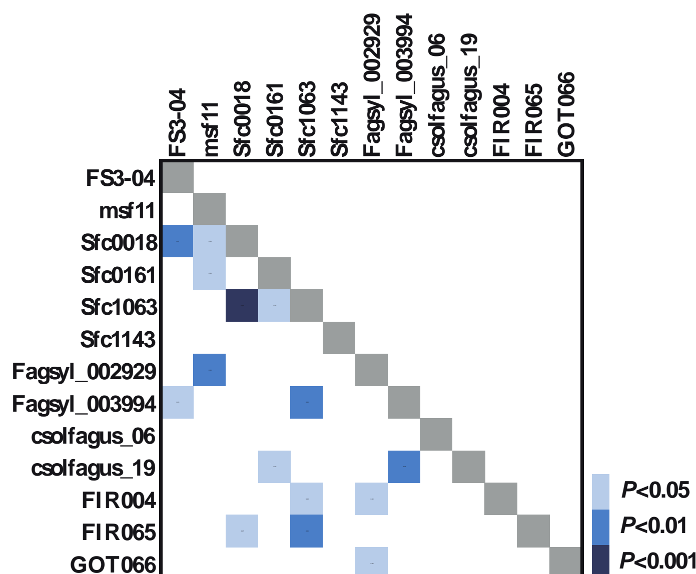


Fig. 2-1 Pairs of SSR loci in linkage disequilibrium (LD) in the saplings

Genetic differentiation and population structure

Genetic differentiation was low but significant for populations of both saplings (G''_{ST} = 0.029; $P < 0.001$) and adults (G''_{ST} = 0.027; $P < 0.001$). The AMOVA analysis revealed that 99% of the variation was within populations, and only 1% between them (Table 2-5). Pairwise population differentiation was also low, ranging between 0.005 and 0.054 for the saplings and between 0 and 0.092 for the adults (Fig. 2-2). Martigny (saplings) and Chamoson (adults) were the most differentiated populations, while Sargans (saplings) and Mels (adults) were the least differentiated. Most of the pairwise G''_{ST} were significant for the saplings (Fig. 2-2).

Table 2-5 AMOVA based on 13 SSR loci for saplings and adults

Source of Variation	d.f.	SS	EV	PV
Saplings				
Among Regions	1	15.73	0.01	0%
Among Populations	10	92	0.04	1%
Within Populations	1498	6321	4.22	99%
Total	1509	6428.73	4.27	100%
Adults				
Among Regions	1	5.44	0	0%
Among Populations	10	63.95	0.04	1%
Within Populations	588	2465.28	4.19	99%
Total	599	2534.67	4.24	100%

d.f. – degrees of freedom, SS – sum of squares, EV- estimated variance, PV – percentage of variation

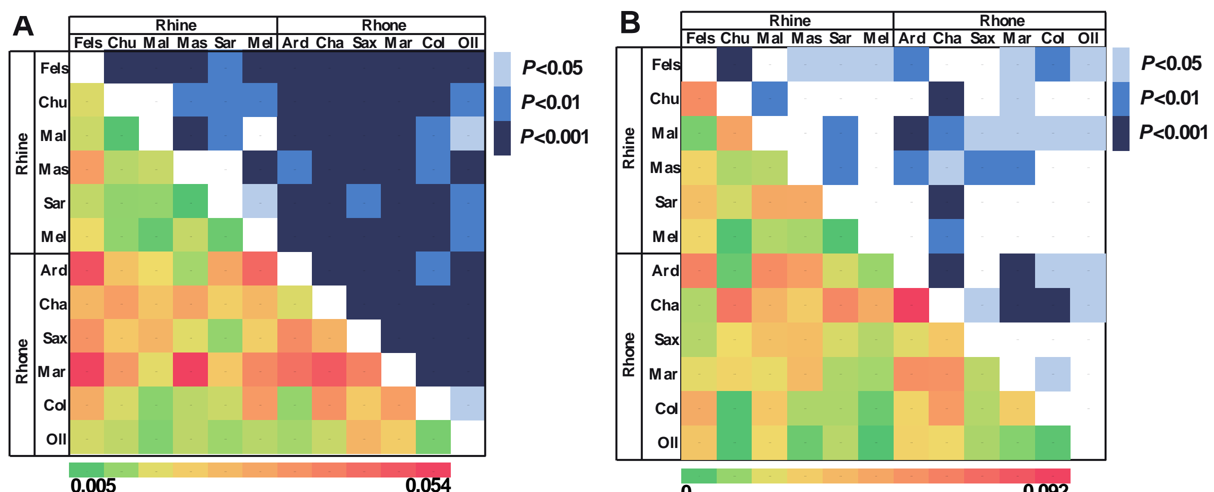


Fig. 2-2 Pairwise genetic differentiation with all 13 SSRs in **A,** saplings and **B,** adults. G''_{ST} values are shown below the diagonal, and P values above the diagonal. Fel - Felsberg; Chu - Chur; Mal - Malans; Mas - Mastrils; Sar - Sargans; Mel - Mels; Ard - Ardon; Cha - Chamoson; Sax - Saxon; Mar - Martigny; Col - Collombey; Oll - Ollon

Analysis of population structure revealed that there are most likely two clusters in the saplings, identifying Chamoson as a genetically different population (Fig. 2-3A). This number of K=2 was supported by both the ΔK method and the highest log probability of the data (Fig. 2-3B). In contrast, in the adults, the structure analysis did not reveal any significant population structure (Fig. 2-4).

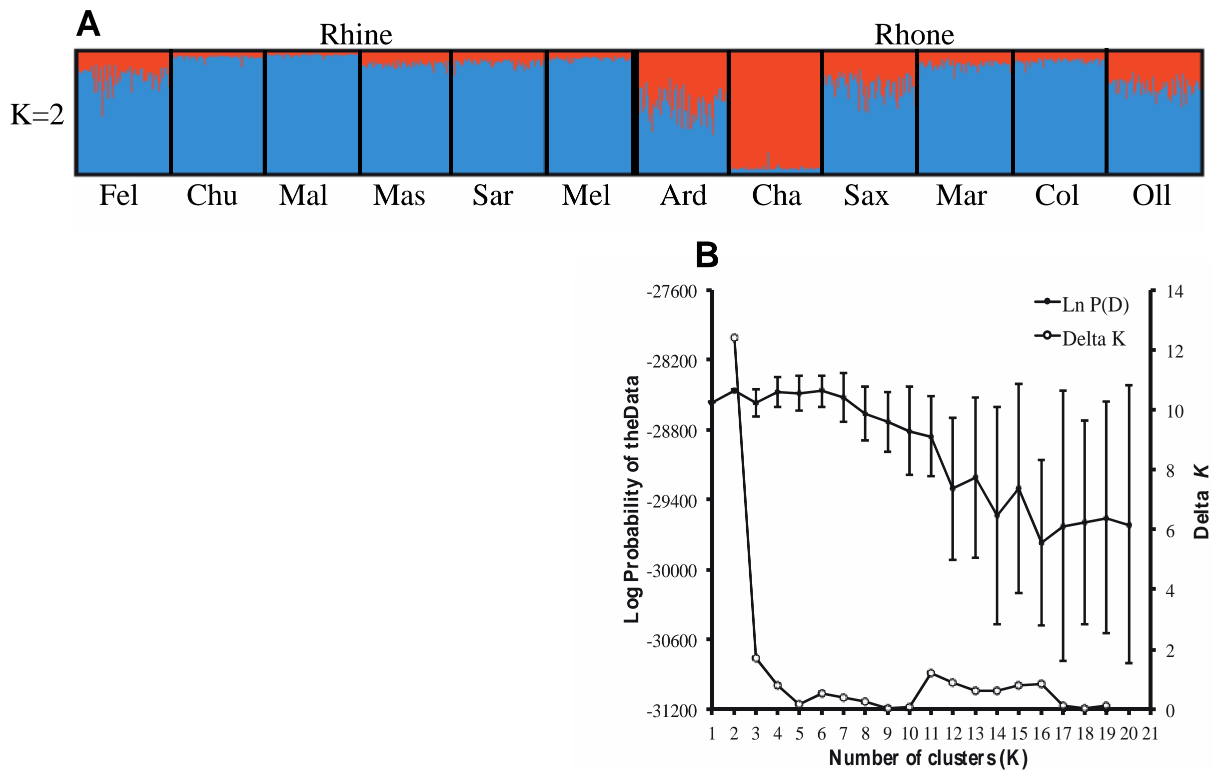


Fig. 2-3 Structure analysis of sapling populations based on the SSR data. **A,** Bar plot indicating the assignment probability of each individual to the two different clusters; **B,** Log probability and ΔK for K=1 to K=20 clusters. Fel - Felsberg; Chu - Chur; Mal - Malans; Mas - Mastrils; Sar - Sargans; Mel - Mels; Ard - Ardon; Cha - Chamoson; Sax - Saxon; Mar - Martigny; Col - Collombey; Oll - Ollon

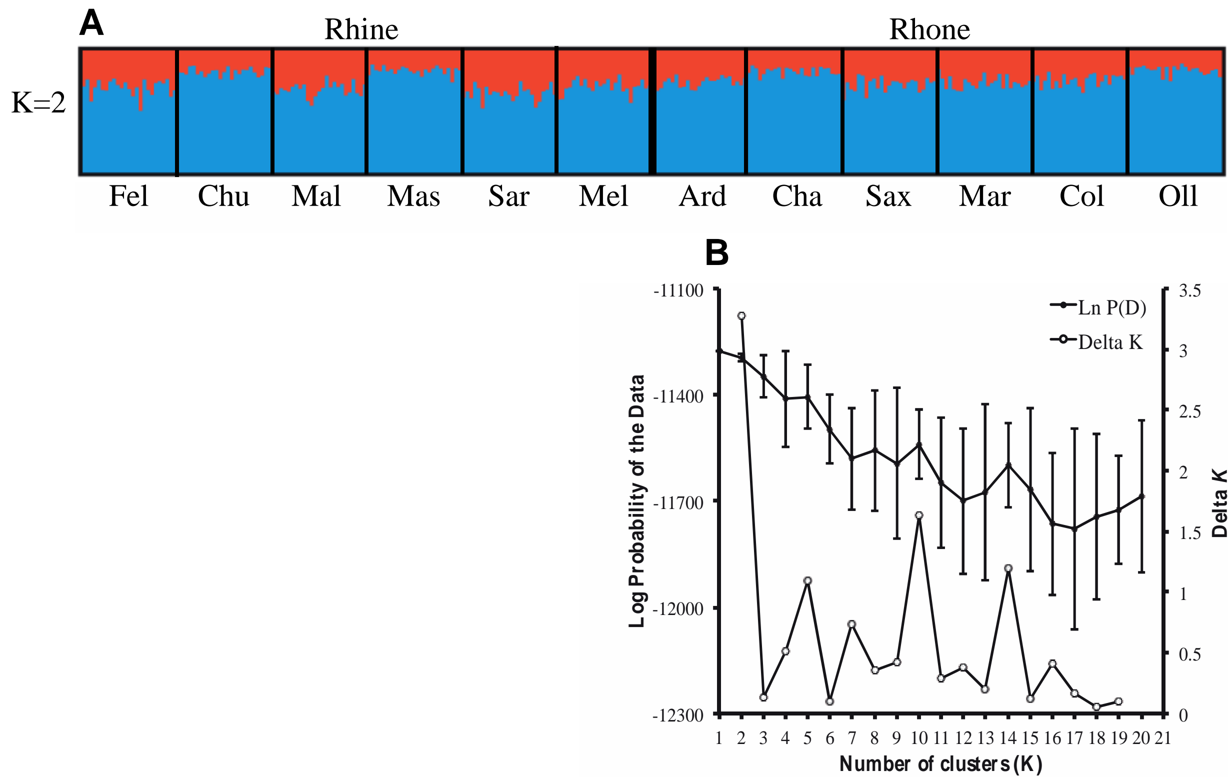


Fig. 2-4 Structure analysis of adult populations based on the SSR data. **A,** Bar plot indicating the assignment probability of each individual to the two different clusters; **B,** Log probability and Δ*K* for *K*=1 to *K*=20 clusters. Fel - Felsberg; Chu - Chur; Mal - Malans; Mas - Mastrils; Sar - Sargans; Mel - Mels; Ard - Ardon; Cha - Chamoson; Sax - Saxon; Mar - Martigny; Col - Collombey; Oll - Ollon.

The dendrograms based on the $(\delta\mu)^2$ genetic distance for the SSR data revealed no significant clustering neither for saplings nor for adults. Tree topology reflected a tendency for populations from the same valley to group together, although the clusters were not strongly supported by the bootstrap values that were mostly very low and less than 60% (Fig. 2-5).

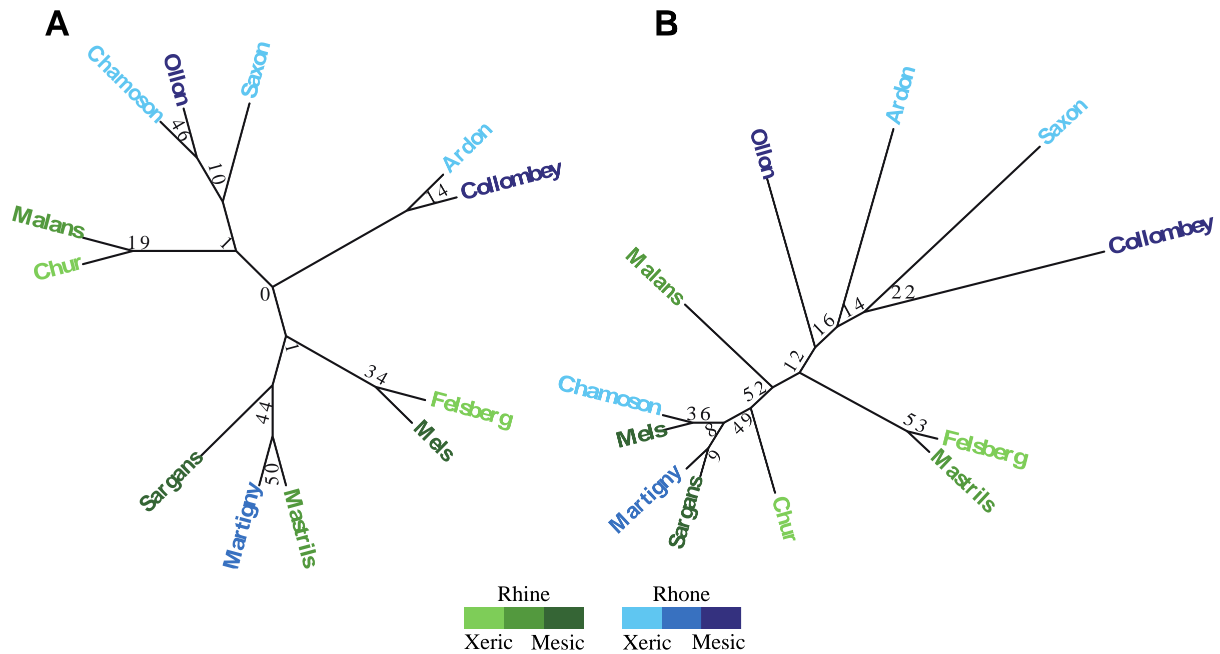


Fig. 2-5 The dendrograms for **A,** saplings and **B,** adults, constructed using the $(\delta\mu)^2$ genetic distance based on SSR data and the neighbor joining clustering method

Outlier analysis

In the saplings, no outlier SSRs were detected by LOSITAN. With Arlequin, one SSR fell outside the 95% confidence interval (GOT066) (Fig. 2-6); however, it did not remain significant after the FDR correction. In contrast, a considerable number of outlier SSR loci were detected by BayeScan: 12 (92%) SSR loci were detected as outliers when doing the analysis with all populations; 11 SSR loci (85%) were detected as outliers in the Rhine valley, and 4 (31%) in the Rhone valley (Fig. 2-7). All outliers detected by BayeScan are possibly under balancing selection.

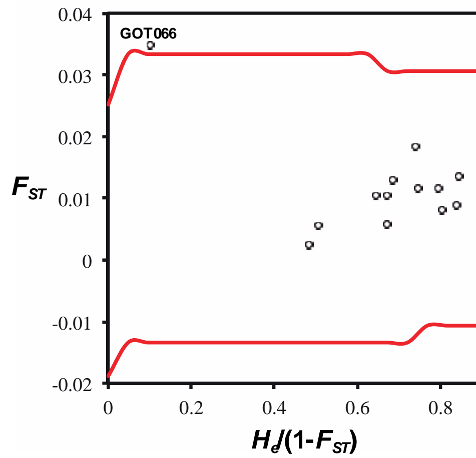


Fig. 2-6 Detection of the outlier SSR loci in the saplings under the hierarchical island model implemented in Arlequin. The red lines represent the 95% confidence interval; the locus falling outside that interval is labeled and potentially under positive selection.

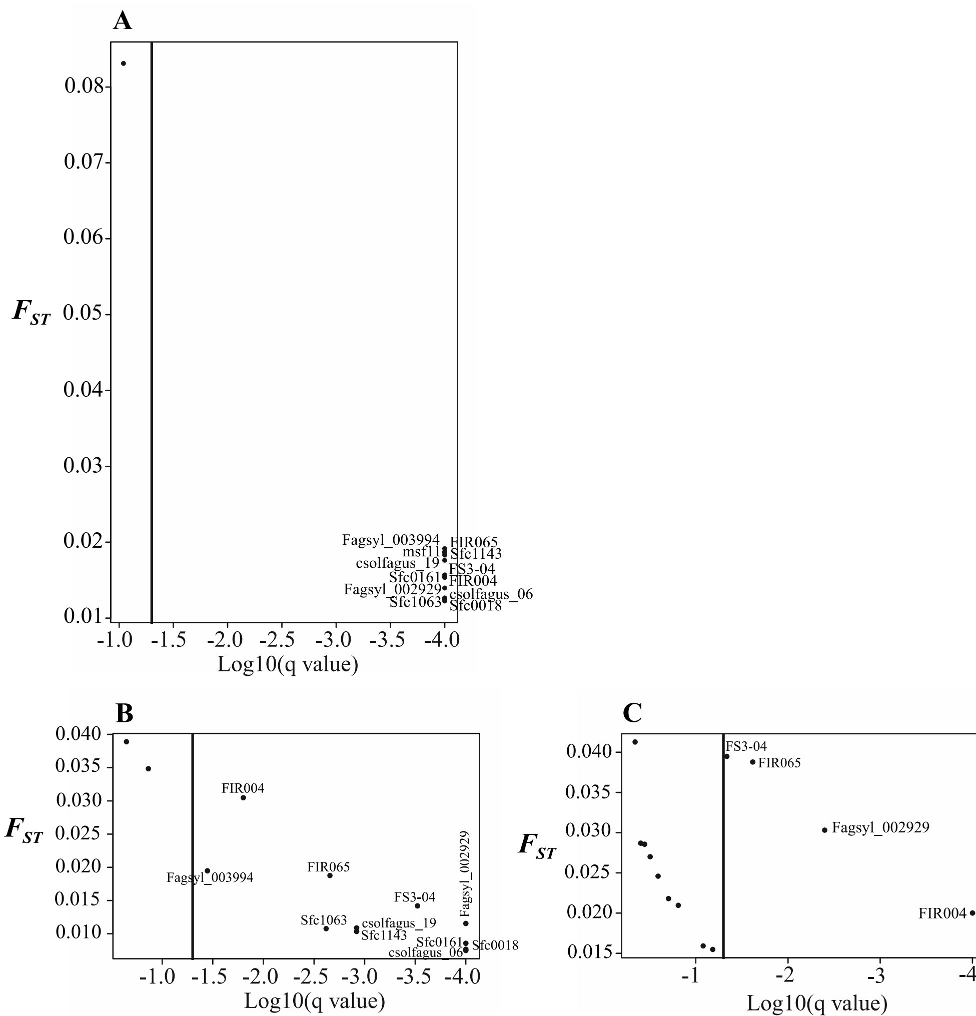


Fig. 2-7 Detection of the outlier SSR loci in the saplings using BayeScan in **A,** all populations; **B,** populations from the Rhine valley; and **C,** populations from the Rhone valley. The vertical line represents the critical *q* value (0.05) used for identifying outlier markers. The labeled markers on the right side are candidates for being under balancing selection

In the adults, the SSR loci detected as outliers by LOSITAN were the same under the stepwise mutation model and the infinite allele model. However, even though the loci FS3-04, FIR065 and csolfagus_06 fell outside the 95% confidence interval when analyzing all populations and populations from each region separately (Fig. 2-8), only the loci FIR065 and csolfagus_06 remained significant after FDR correction in the analysis using all populations. These two loci (15%) are likely under balancing selection - they have low F_{ST}, but high H_e values (Fig. 2-8). Arlequin identified the FS3-04 locus as an outlier as well (Fig. 2-9); however, it did not remain significant after FDR correction. BayeScan identified 12 (92%) outlier SSR loci when doing the analysis with all populations, 7 (54%) in the analysis with populations from the Rhine valley, and 4 (31%) in the analysis with populations from the Rhone valley (Fig. 2-10). All outliers detected by BayeScan are likely under balancing selection.

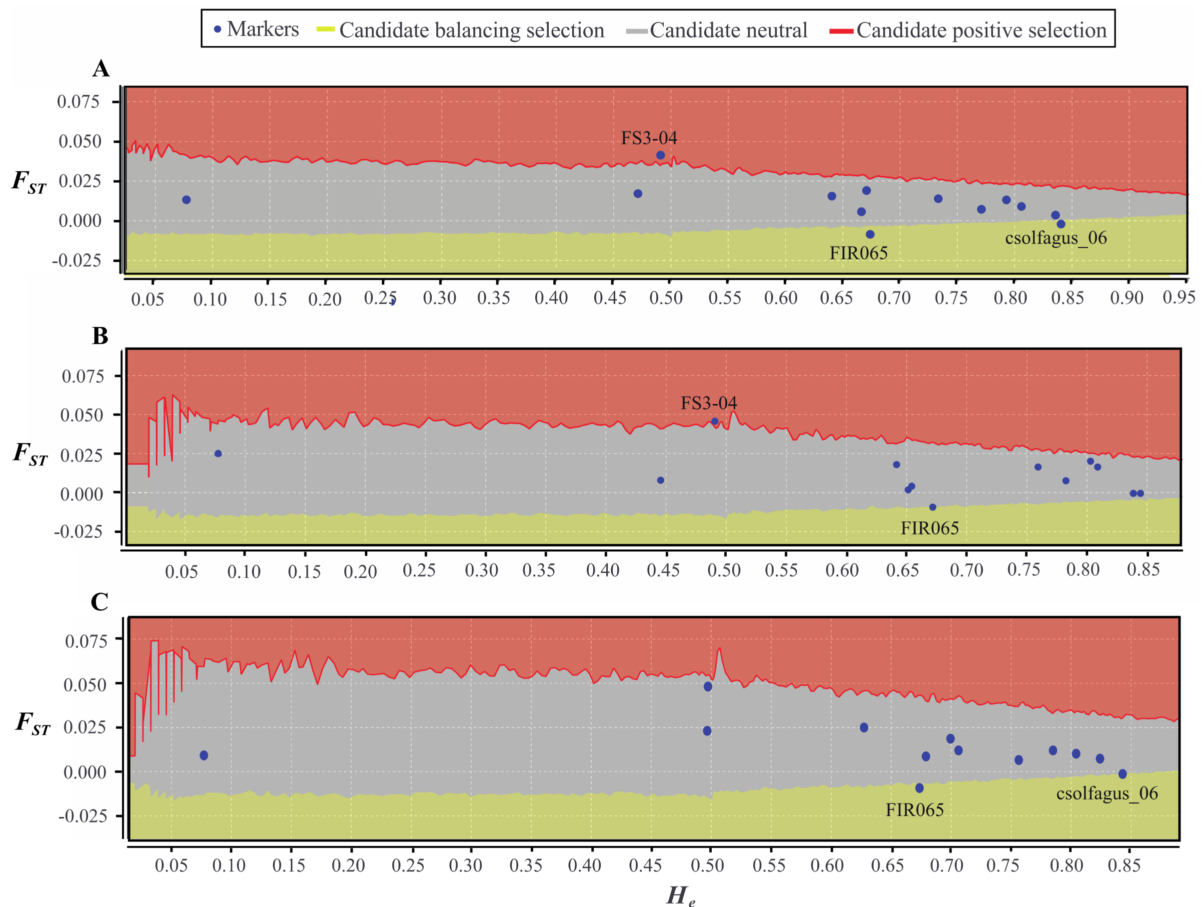


Fig. 2-8 Outlier SSR loci detected in adults using LOSITAN under an island model. Results show the distribution of observed F_{ST} values for each SSR marker along their mean within population heterozygosities (H_e) in **A,** all populations; **B,** populations from the Rhine valley; and **C,** populations from the Rhone valley. The gray area represents the 95% confidence interval; loci falling outside that interval are labeled and could be potentially under selection.

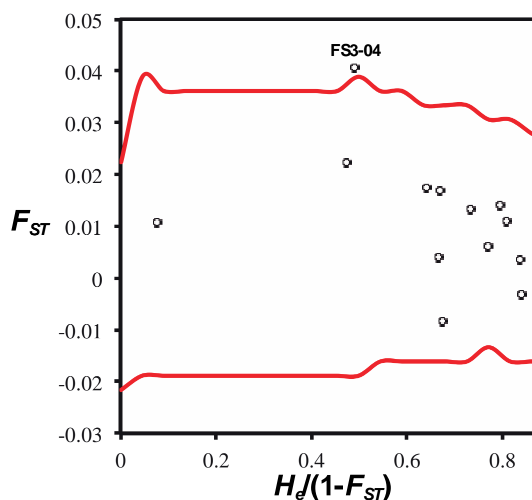


Fig. 2-9 Detection of the outlier SSR loci in the adults under a hierarchical island model implemented in Arlequin. The red lines delineate the 95% confidence interval; the locus falling outside that interval is labeled and potentially under positive selection.

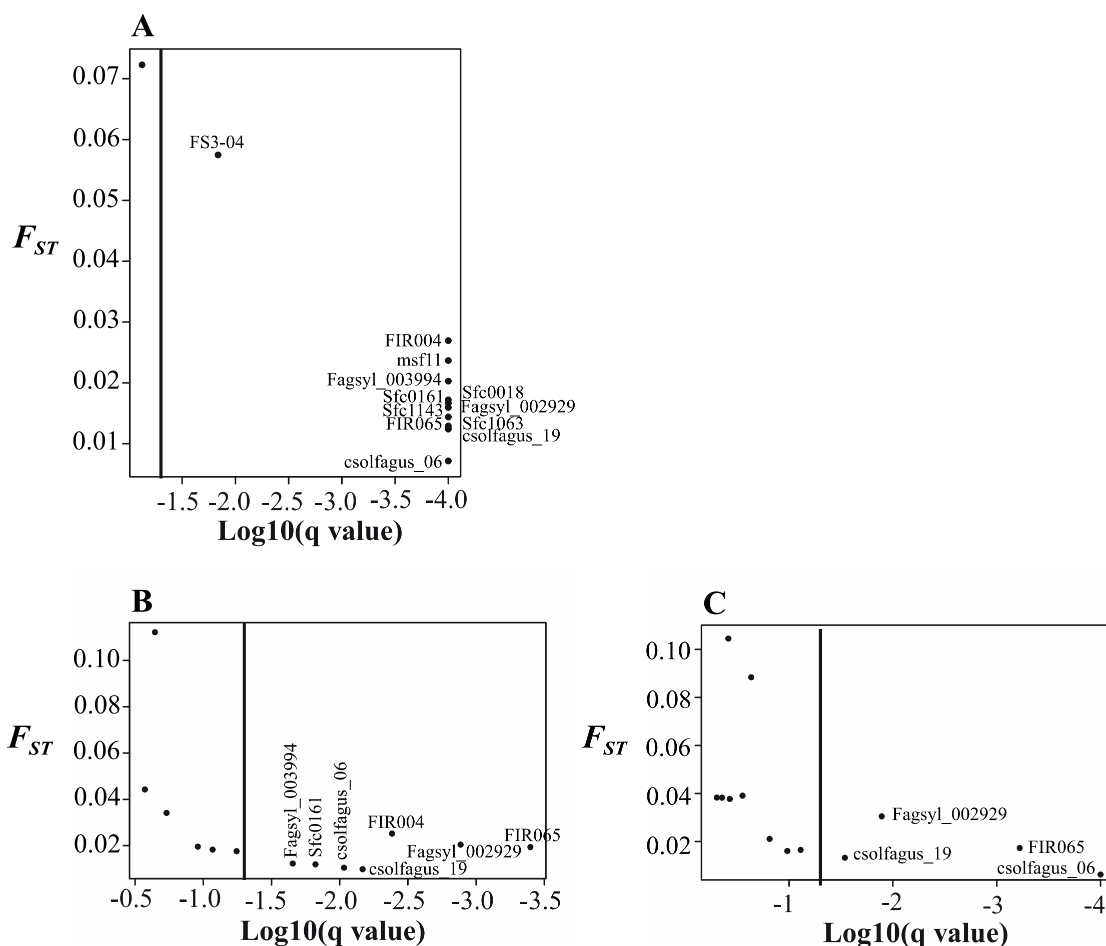


Fig. 2-10 Detection of the outlier SSR loci in the adults using BayeScan across **A,** all populations; **B,** populations from the Rhine valley; and **C,** populations from the Rhone valley. The vertical line represents the critical q value (0.05) used for identifying outlier markers. The labeled markers on the right side are likely under balancing selection

Among the outlier SSR loci detected in adult populations, the FS3-04, csolfagus_06 and FIR065 loci are very likely under selection, because they were detected by at least two of the outlier methods (Table 2-6). The rest of the outliers in adults and saplings could be false.

Table 2-6 List of the SSR loci detected as outliers by at least one of the three methods in saplings and adults

Saplings				Adults			
LOSITAN	**ARLEQUIN**	**BAYESCAN**	**Selection**	**LOSITAN**	**ARLEQUIN**	**BAYESCAN**	**Selection**
		FS3-04*	Balancing	**FS3-04**	**FS3-04**	**FS3-04***	Positive
		msf11*	Balancing			msf11*	Balancing
		Sfc0018*	Balancing			Sfc0018*	Balancing
		Sfc0161*	Balancing			Sfc0161*	Balancing
		Sfc1063*	Balancing			Sfc1063*	Balancing
		Sfc1143*	Balancing			Sfc1143*	Balancing
None		Fagsyl_002929*	Balancing			Fagsyl_002929*	Balancing
		Fagsyl_003994*	Balancing			Fagsyl_003994*	Balancing
		csolfagus_06*	Balancing	**csolfagus_06***		**csolfagus_06***	Balancing
		csolfagus_19*	Balancing			csolfagus_19*	Balancing
		FIR004*	Balancing			FIR004*	Balancing
		FIR065*	Balancing	**FIR065***		**FIR065***	Balancing
	GOT066		Positive				

* Loci that remained significant after correcting for FDR. Loci highlighted by bold font were detected by two or all three methods

The allele 201 in the FS3-04 locus was under positive selection and had a higher frequency in some of the adult populations with low and intermediate level of precipitation in both valleys compared to populations with high level of precipitation (Fig. 2-11A). The outlier locus csolfagus_06 was likely under balancing selection, which is in consensus with an excess of heterozygosity observed in some populations, especially in Chamoson (Fig. 2-11B). Likewise, at the FIR065 locus, also under balancing selection, an excess of observed heterozygosity was observed in most of the populations, especially in Malans, Chamoson and Collombey, located in different valleys and having different levels of precipitation (Fig.2-11C).

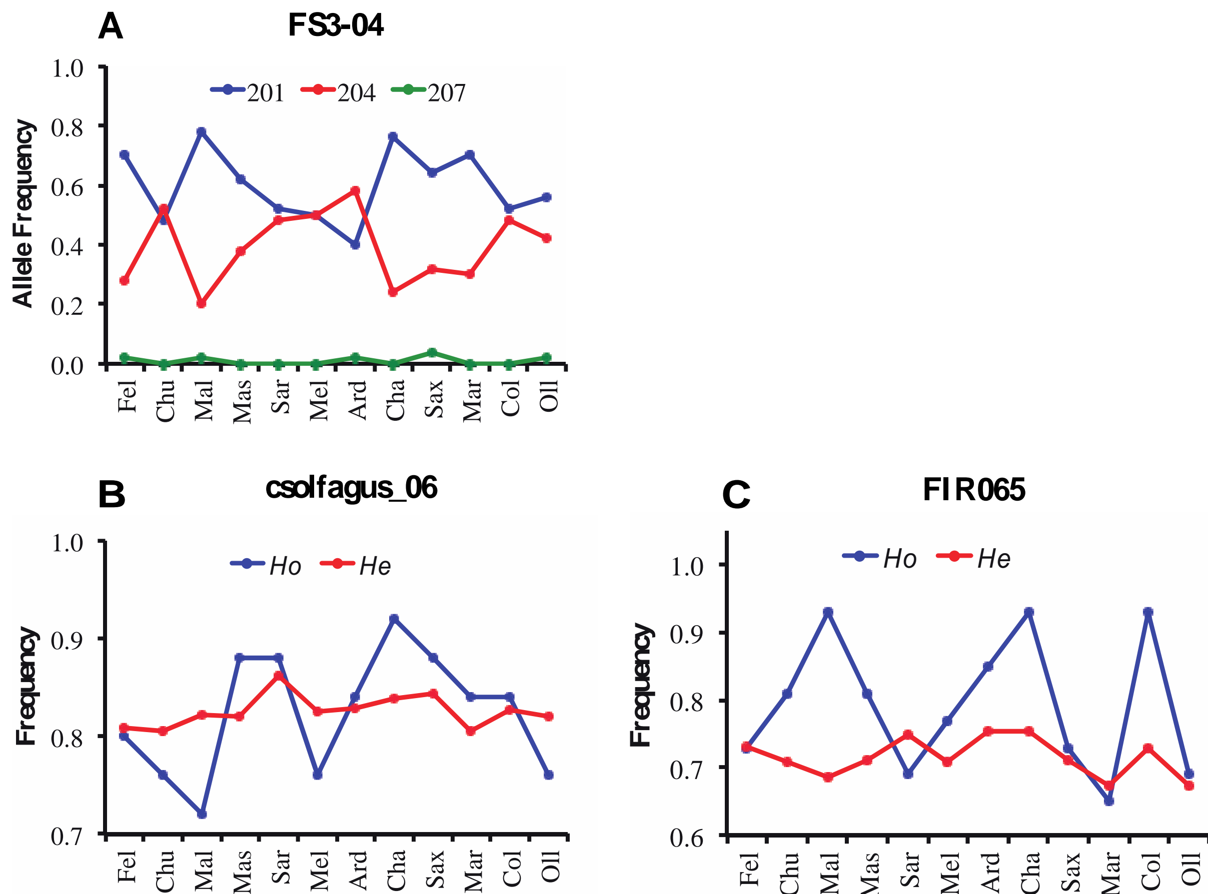


Fig. 2-11 Allele and heterozygosity frequency distribution in adults for the loci under **A,** positive selection; **B** and **C,** balancing selection. H_o - observed heterozygosity; H_e - expected heterozygosity

SNPs

Genetic variability and linkage disequilibrium (LD)

From the 76 selected SNPs for genotyping, 6 were monomorphic (APX1_2, PhyB, 50_320, 52_1_249, 92_166, 110_1_111); thus, the remaining 70 were used for the analysis. The level of genetic diversity varied among SNP loci. The observed and expected heterozygosities were the lowest for the SNP ALDH_4 in both saplings ($H_o = 0.016$, $H_e = 0.018$) and adults ($H_o = 0.017$, $H_e = 0.016$). The SNP CP10_377 had the highest observed heterozygosity in the saplings ($H_o = 0.519$), but the SNP 17_1081 in the adults ($H_o = 0.808$), while the highest expected heterozygosity was found in the SNP 92_630 in the saplings ($H_e = 0.498$) and in the SNP PP2C_1200 in the adults ($H_e = 0.502$) (Table 2-7). In general, the fixation index F_{IS} was close to zero, and no significant deviations from Hardy-Weinberg equilibrium were found, except for the SNPs CP10_749, 50_232, 91_2_479 and 92_630 that showed heterozygous deficit in the saplings, and for the SNPs 17_1081 and 92_630 that showed excess and deficit

of heterozygotes, respectively, in the adults (Table 2-7). Analysis of genetic diversity revealed no significant differences between adults and saplings.

Table 2-7 Diversity parameters and fixation index for SNP loci over all populations of saplings and adults

Locus	Saplings			Adults		
	H_o	H_e	F_{IS}	H_o	H_e	F_{IS}
ALDH_1	0.283	0.297	0.037	0.323	0.308	-0.071
ALDH_2	0.306	0.318	0.031	0.363	0.336	-0.104
ALDH_3	0.033	0.035	0.048	0.05	0.048	-0.058
ALDH_4	0.016	0.018	0.127	0.017	0.016	-0.037
IDH_1	0.481	0.49	0.01	0.483	0.478	-0.031
IDH_3	0.291	0.32	0.083	0.286	0.297	0.018
IDH_4	0.482	0.491	0.01	0.475	0.478	-0.014
APX1_1	0.227	0.227	-0.004	0.291	0.277	-0.071
APX4_1	0.225	0.228	0.007	0.211	0.216	0.006
APX4_2	0.495	0.491	-0.016	0.49	0.483	-0.034
ERD	0.463	0.476	0.021	0.457	0.49	0.049
Dhn_1	0.41	0.422	0.022	0.5	0.451	-0.13
Dhn_2	0.174	0.179	0.022	0.193	0.188	-0.045
GPX	0.13	0.126	-0.04	0.113	0.112	-0.029
CP10_65	0.318	0.313	-0.023	0.327	0.291	-0.144
CP10_67	0.075	0.077	0.018	0.063	0.062	-0.05
CP10_377	0.519	0.495	-0.057	0.525	0.5	-0.073
CP10_442	0.406	0.393	-0.044	0.441	0.4	-0.124
CP10_503	0.143	0.131	-0.101	0.111	0.106	-0.07
CP10_749	0.297	0.325	0.080*	0.3	0.307	0.003
CP10_1317	0.261	0.265	0.006	0.287	0.256	-0.144
CP10_1428	0.408	0.393	-0.047	0.444	0.396	-0.145
CysPro_118	0.292	0.311	0.051	0.294	0.318	0.054
CysPro_202	0.037	0.035	-0.053	0.03	0.035	0.112
CysPro_728	0.205	0.211	0.022	0.199	0.2	-0.012
CysPro_783	0.294	0.312	0.051	0.295	0.318	0.055
DAG_81	0.186	0.179	-0.043	0.2	0.193	-0.055
DAG_289	0.184	0.179	-0.039	0.201	0.194	-0.054
DAG_1059	0.136	0.15	0.081	0.148	0.136	-0.115
His3C1_292	0.382	0.391	0.015	0.424	0.439	0.016
His3C2_104	0.053	0.054	0.014	0.047	0.046	-0.042
His3C2_186	0.189	0.184	-0.033	0.167	0.168	-0.017
His3C2_260	0.113	0.111	-0.031	0.094	0.096	0
NAC_854	0.295	0.311	0.042	0.316	0.318	-0.016
NAC_962	0.161	0.163	0.002	0.203	0.191	-0.083
NAC_1300	0.44	0.458	0.032	0.448	0.444	-0.029
PP2C_315	0.035	0.036	0.027	0.023	0.023	-0.026
PP2C_391	0.477	0.476	-0.009	0.44	0.484	0.073
PP2C_791	0.061	0.062	0.001	0.05	0.048	-0.073

Locus	Saplings			Adults		
	H_o	H_e	F_{IS}	H_o	H_e	F_{IS}
PP2C_941	0.494	0.492	-0.012	0.47	0.501	0.042
PP2C_1200	0.494	0.493	-0.011	0.47	0.502	0.044
7_258	0.151	0.152	-0.006	0.156	0.155	-0.03
7_520	0.073	0.072	-0.024	0.05	0.048	-0.055
17_880	0.046	0.052	0.107	0.067	0.077	0.111
17_1081	0.461	0.491	0.053	0.808	0.488	-0.689***
39_256	0.473	0.489	0.025	0.436	0.488	0.088
39_282	0.472	0.49	0.03	0.465	0.493	0.038
50_39	0.285	0.295	0.028	0.295	0.31	0.032
50_232	0.475	0.482	0.007*	0.49	0.498	-0.004
52_1_235	0.386	0.382	-0.019	0.364	0.386	0.038
52_1_368	0.39	0.385	-0.022	0.363	0.388	0.044
68_277	0.489	0.454	-0.086	0.465	0.461	-0.028
68_313	0.478	0.45	-0.071	0.455	0.457	-0.016
88_1_450	0.044	0.048	0.072	0.037	0.036	-0.032
88_1_727	0.299	0.294	-0.026	0.308	0.294	-0.067
88_1_803	0.273	0.277	0.006	0.284	0.277	-0.045
91_2_57	0.394	0.396	-0.002	0.428	0.407	-0.074
91_2_141	0.4	0.428	0.056	0.435	0.423	-0.052
91_2_231	0.411	0.424	0.022	0.44	0.425	-0.057
91_2_448	0.413	0.425	0.021	0.441	0.426	-0.056
91_2_479	0.384	0.417	0.071*	0.418	0.418	-0.021
91_2_504	0.412	0.425	0.024	0.44	0.426	-0.054
92_352	0.173	0.172	-0.013	0.188	0.18	-0.065
92_630	0.287	0.498	0.419***	0.242	0.499	0.504***
110_1_293	0.433	0.472	0.074	0.413	0.459	0.081
110_1_423	0.465	0.479	0.021	0.484	0.475	-0.039
110_1_450	0.468	0.479	0.014	0.478	0.475	-0.026
154_2_137	0.484	0.476	-0.024	0.476	0.49	0.008
154_2_371	0.394	0.394	-0.008	0.405	0.442	0.065
154_2_617	0.192	0.183	-0.058	0.185	0.167	-0.129
Mean	**0.301**	**0.309**	**0.013**	**0.311**	**0.310**	**-0.026**

H_o – observed heterozygosity, H_e – expected heterozygosity, F_{IS} – fixation index, *$P<0.05$, **$P<0.01$, ***$P<0.001$

Only slight differences in heterozygosity were observed between populations (Table 2-8). Expected heterozygosity ranged between 0.298 (Ardon) and 0.325 (Mastrils) in the saplings, and between 0.288 (Ardon) and 0.335 (Sargans) in the adults (Table 2-8). In general, no significant deviations from Hardy-Weinberg equilibrium were found, except for the Mastrils, Sargans and Ollon populations in the saplings, and the population Mastrils in the adults. Likewise, there were no significant differences between the two regions, neither in the saplings nor the adults.

Table 2-8 Diversity parameters at SNP loci for saplings and adult populations

Population	Saplings			Adults		
	H_o	H_e	F_{IS}	H_o	H_e	F_{IS}
Rhine						
Felsberg	0.314	0.317	0.025	0.296	0.301	-0.006
Chur	0.308	0.313	0.015	0.325	0.315	-0.061
Malans	0.296	0.308	0.023	0.308	0.317	-0.010
Mastrils	0.294	0.325	0.064*	0.337	0.323	-0.055*
Sargans	0.318	0.314	-0.024*	0.343	0.335	-0.032
Mels	0.286	0.306	0.043	0.332	0.307	-0.090
Mean	0.303	0.314	0.027	0.324	0.316	-0.044
Rhone						
Ardon	0.307	0.298	-0.037	0.314	0.288	-0.095
Chamoson	0.304	0.311	0.024	0.319	0.319	-0.024
Saxon	0.282	0.299	0.036	0.293	0.313	0.017
Martigny	0.292	0.301	0.015	0.313	0.320	0.002
Collombey	0.302	0.306	-0.002	0.291	0.298	-0.011
Ollon	0.309	0.306	-0.019*	0.263	0.288	0.061
Mean	0.299	0.304	0.007	0.299	0.304	-0.002
Grand mean	0.301	0.309	0.013	0.311	0.310	-0.026

N – sample size, H_o – observed heterozygosity, H_e – expected heterozygosity, F_{IS} – fixation index, *P<0.05

In both saplings and adults, LD was mainly found between SNPs belonging to the same gene (Figs. 2-12 and 2-13). In the saplings, significant LD was observed for 134 pairs (5.5%) of all the possible pair combinations of SNPs (2415), and 68 of them were found between SNPs from the same gene (Fig. 2-12). Similarly, for populations of adults, 107 pairs (4.4%) of all the possible pairs showed significant LD, and 59 of them were found between SNPs from the same gene (Fig. 2-13).

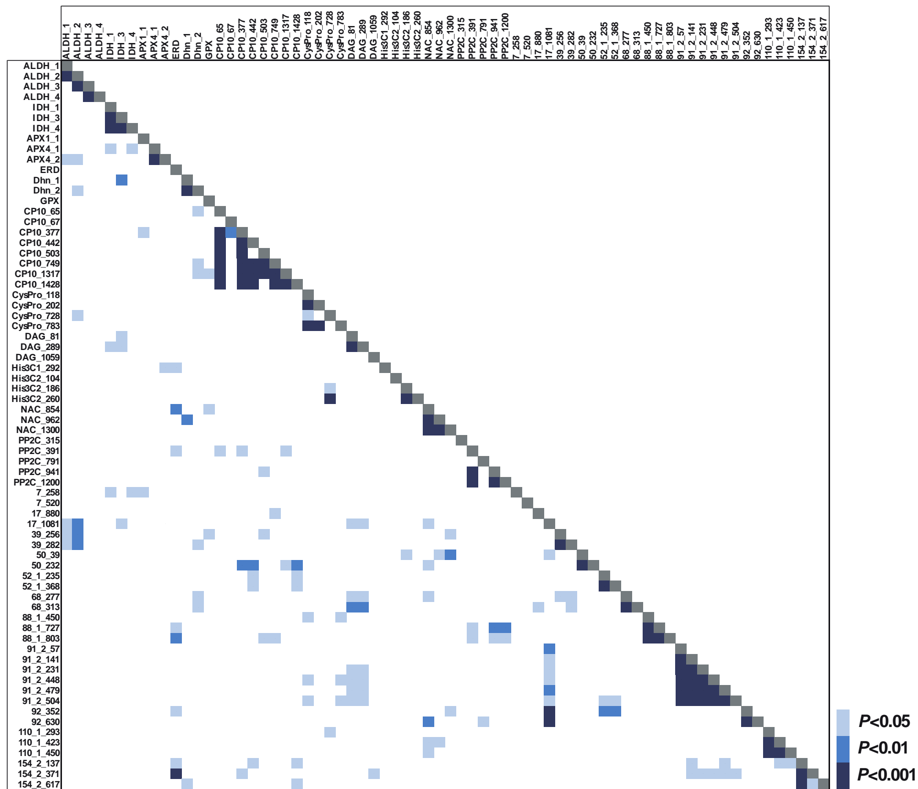


Fig. 2-12 Linkage disequilibrium (LD) plot between pairs of SNPs over all populations of saplings

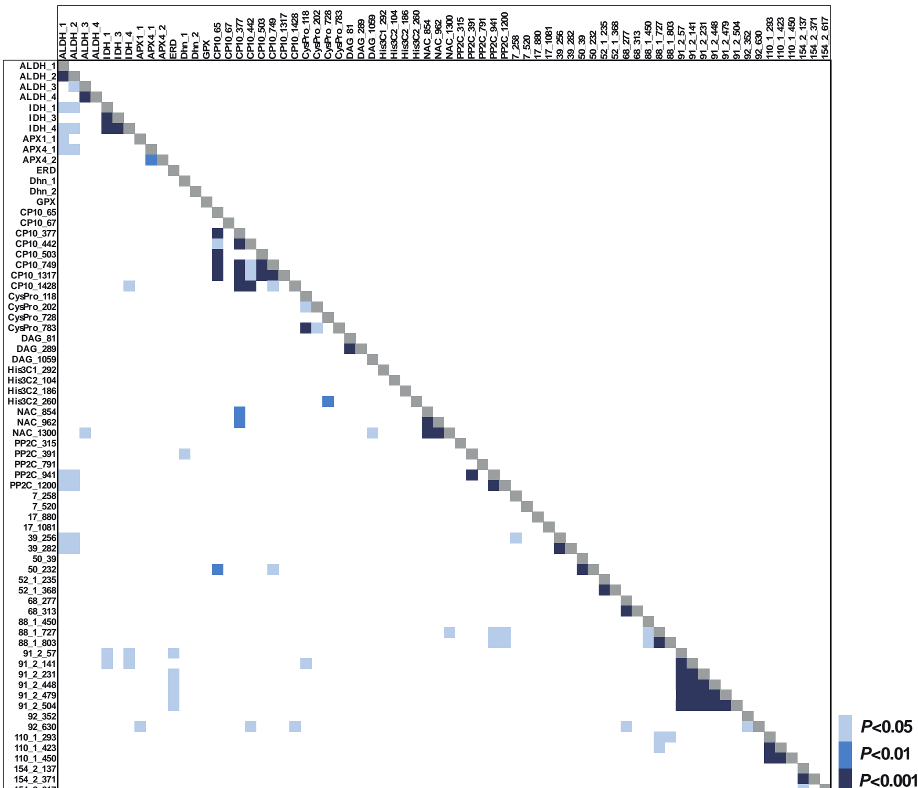


Fig. 2-13 Linkage disequilibrium (LD) plot between pairs of SNPs over all populations of adults

Genetic differentiation and population structure

Genetic differentiation was low but significant for populations of both saplings (G''_{ST} = 0.020; $P < 0.001$) and adults (G''_{ST} = 0.016; $P < 0.001$). AMOVA analysis revealed that 98-99% of the variation was within populations, and only 1-2% among them (Table 2-9). Pairwise population differentiation was also low, ranging between 0 and 0.044 for the saplings and between 0 and 0.062 for the adults (Fig. 2-14). Chamoson was the population with the highest pairwise differentiation in the saplings, while Sargans and Malans had the lowest differentiation. In the adults, Chamoson was also the population with the highest pairwise differentiation, followed by Ardon; while Martigny and Collombey were the least differentiated (Fig. 2-14).

Table 2-9 AMOVA based on SNPs for saplings and adults

Source of Variation	d.f.	SS	EV	PV
Saplings				
Among Regions	1	67.39	0.05	0%
Among Populations	10	275.29	0.13	2%
Within Populations	1498	16644.39	11.11	98%
Total	1509	16987.06	11.29	100%
Adults				
Among Regions	1	31.24	0.05	0%
Among Populations	10	159.49	0.10	1%
Within Populations	588	6553.16	11.14	99%
Total	599	6743.90	11.29	100%

d.f. – degrees of freedom, SS – sum of squares, EV- estimated variance, PV – percentage of variation

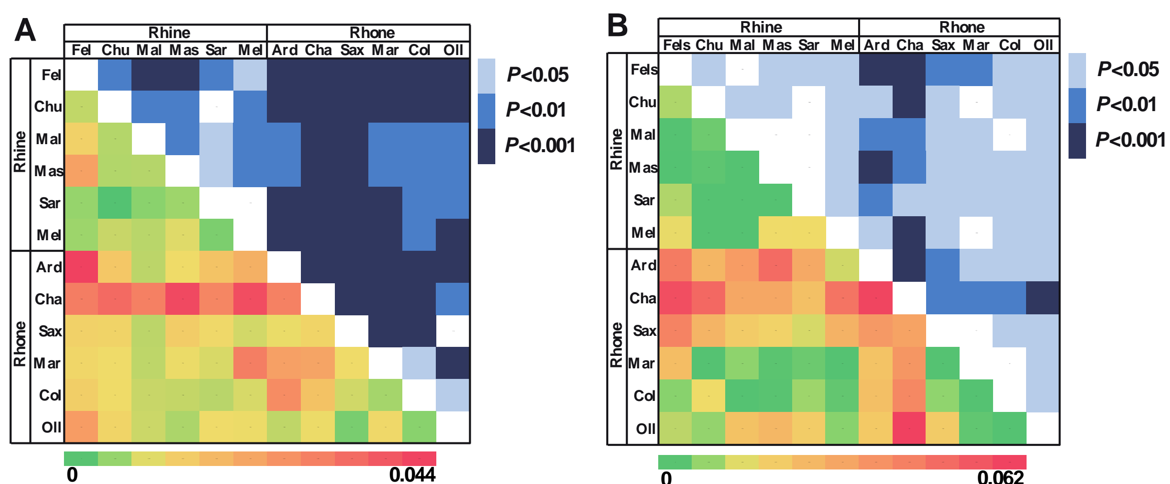


Fig. 2-14 Pairwise genetic differentiation estimated with all SNPs in **A,** saplings; and **B,** adults. G''_{ST} values are shown below the diagonal, and P values above the diagonal. Fel - Felsberg; Chu - Chur; Mal - Malans; Mas - Mastrils; Sar - Sargans; Mel - Mels; Ard - Ardon; Cha - Chamoson; Sax - Saxon; Mar - Martigny; Col - Collombey; Oll - Ollon.

Analysis of population structure using SNP markers revealed that there is a weak population structure in both saplings and adults (Figs. 2-15 and 2-16). Although, based on the ΔK method there could be two clusters (K = 2), the bar plots do not support it much.

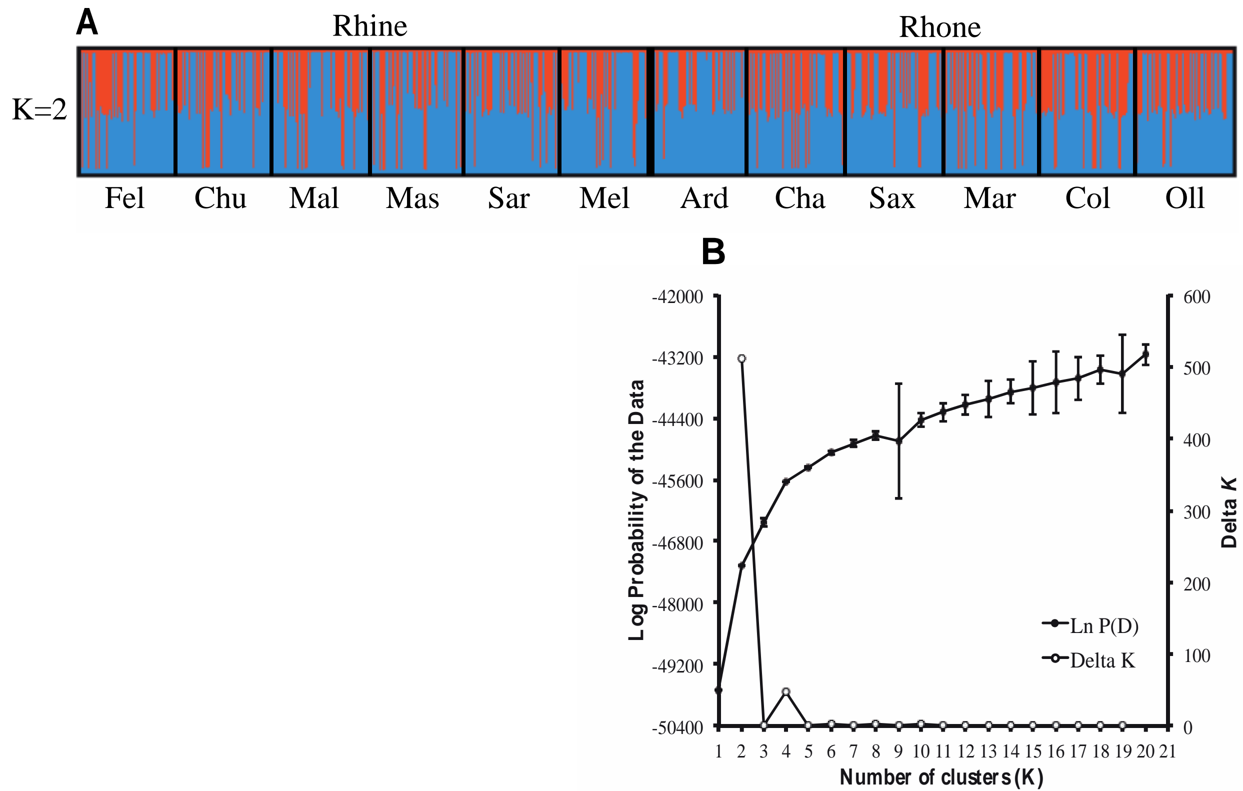


Fig. 2-15 Population structure of populations of saplings based on SNP data. **A,** Bar plot indicating the assignment probability of each individual to the two possible clusters; **B,** Log probability and ΔK for K = 1 to K = 20 clusters. Fel - Felsberg; Chu - Chur; Mal - Malans; Mas - Mastrils; Sar - Sargans; Mel - Mels; Ard - Ardon; Cha - Chamoson; Sax - Saxon; Mar - Martigny; Col - Collombey; Oll - Ollon.

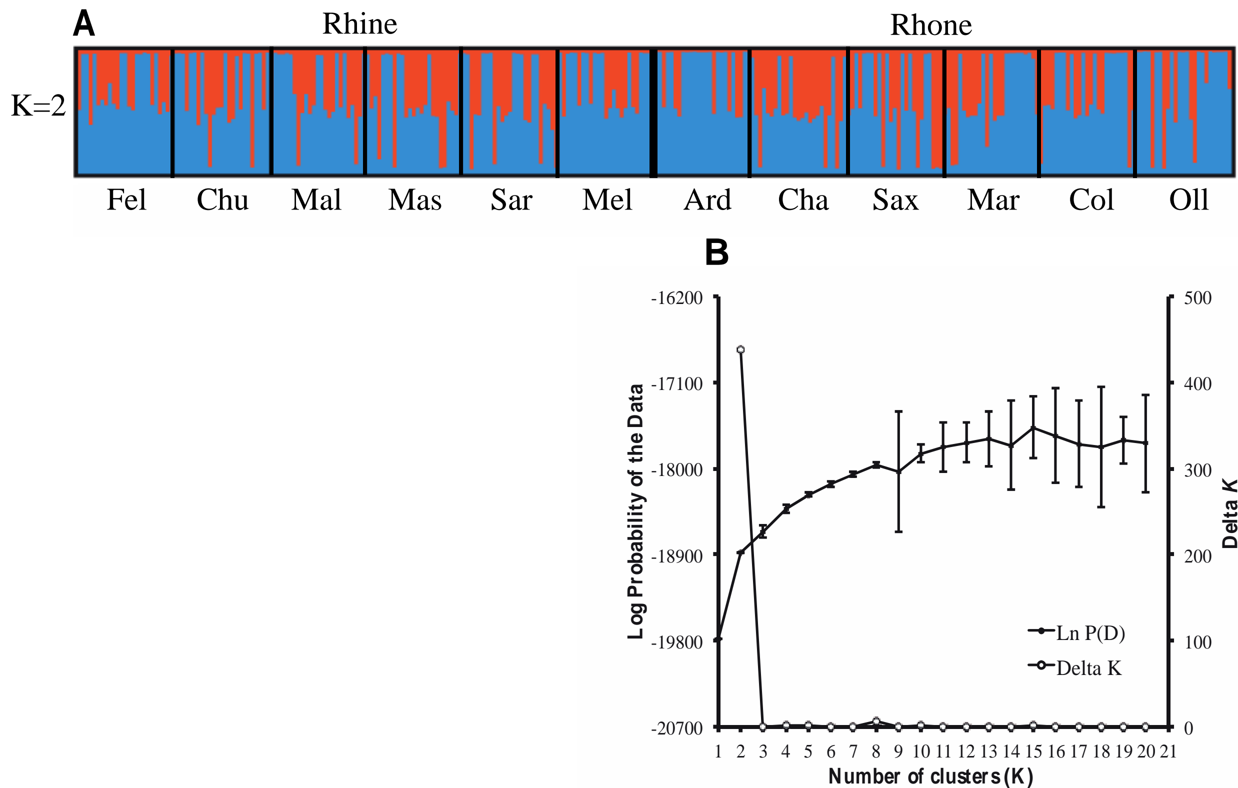


Fig. 2-16 Population structure of populations of adults based on SNP data. **A,** Bar plot indicating the assignment probability of each individual to the two possible clusters; **B,** Log probability and ΔK for $K = 1$ to $K = 20$ clusters. Fel - Felsberg; Chu - Chur; Mal - Malans; Mas - Mastrils; Sar - Sargans; Mel - Mels; Ard - Ardon; Cha - Chamoson; Sax - Saxon; Mar - Martigny; Col - Collombey; Oll - Ollon

Although neighbor joining trees based on the Nei's standard genetic distance (1972) revealed a tendency of the populations from the same valley to cluster together; however, the clusters were not well supported by the bootstrap values that were mostly very low and less than 60% (Fig. 2-17). Two pairs of sapling populations, Sargans and Chur and Felsberg and Mels (Fig. 2-17A) had high bootstrap support, but they were not from the neighboring locations and occurred in different environments.

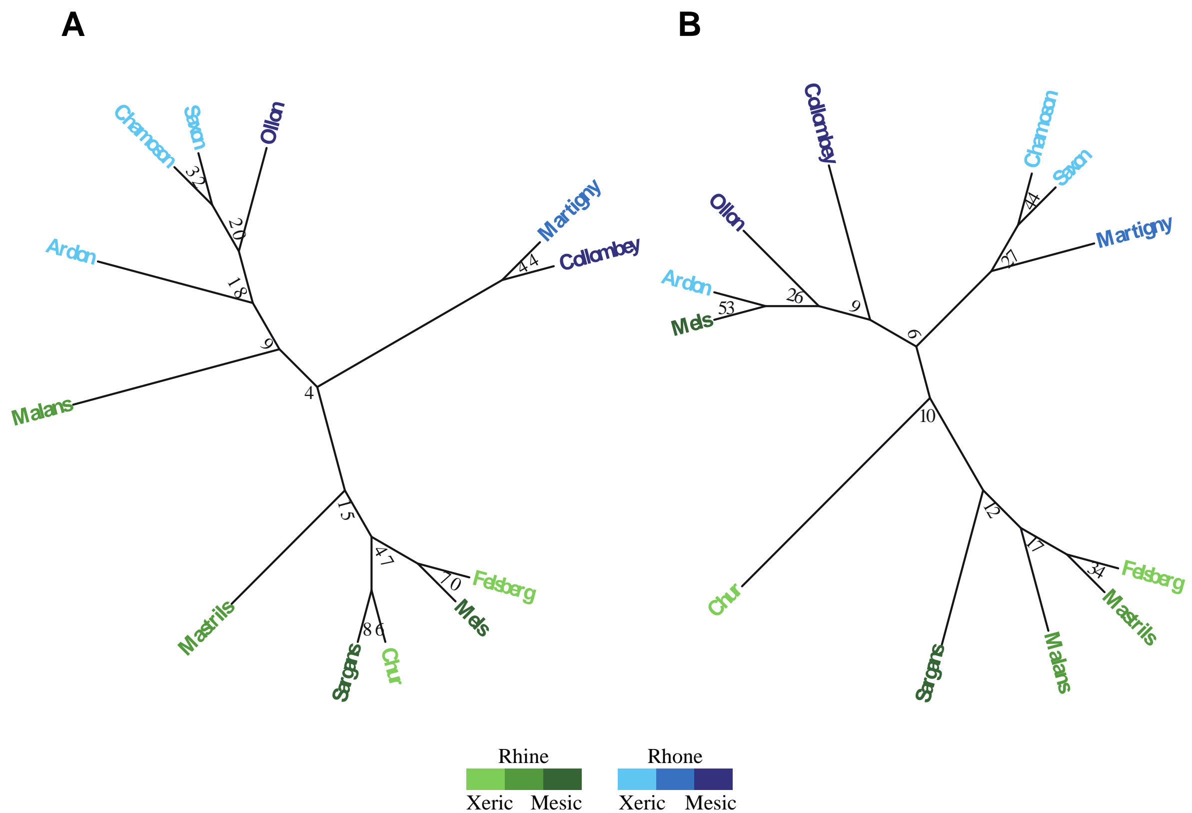


Fig. 2-17 Neighbor joining dendrograms using Nei's standard genetic distance (1972) based on SNP data for **A**, saplings; and **B,** adults.

Outlier Analysis

In the saplings, the outlier SNP ALDH_4 was detected by LOSITAN as being under balancing selection in populations from the Rhone valley (Fig. 2-18). This locus remained significant after correction for multiple testing. No outliers were identified by LOSITAN when doing the analysis with all populations together and with populations from the Rhine valley. Arlequin identified the SNPs ERD, CysPro_202 and NAC_962 as outliers that are likely under positive selection (Fig. 2-19). However, none of them remained significant after the FDR correction. No outlier SNP loci were identified by BayeScan.

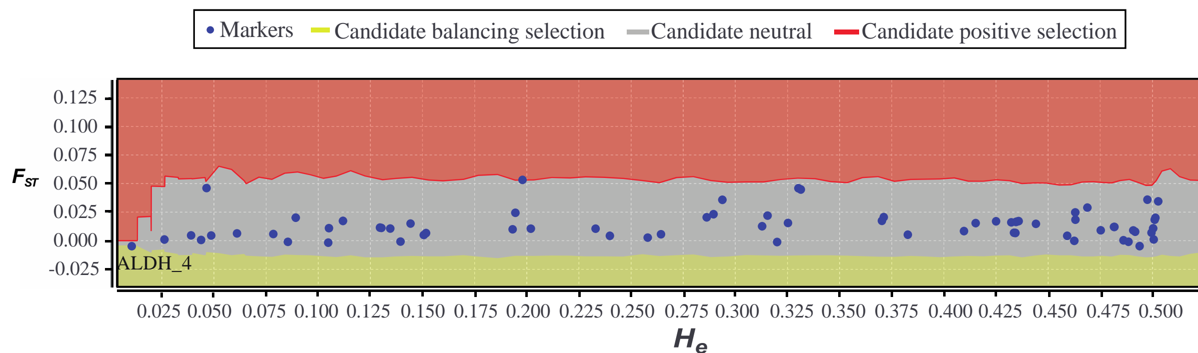


Fig. 2-18 Outlier SNPs detected in saplings from the Rhone valley with LOSITAN, under an island model. Results show the distribution of observed F_{ST} values for each SNP along their mean within population heterozygosities (H_e). The gray area represents the 95% confidence interval; loci falling outside that interval are labeled and could be potentially under selection.

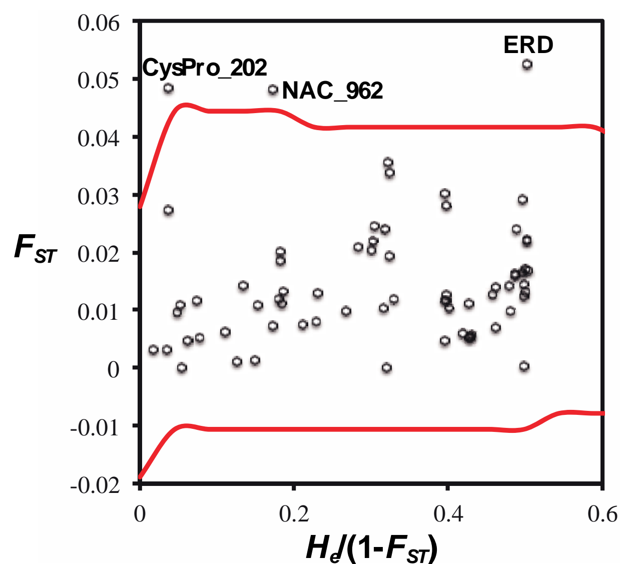


Fig. 2-19 Detection of outlier SNP loci in the saplings using the hierarchical island model implemented in Arlequin. The red lines delineate the 95% confidence interval; loci falling outside that interval are labeled and could be potentially under positive selection.

Compared to the saplings, more outlier SNPs were identified in the adults. In the LOSITAN analysis, 15 SNPs fell outside the 95% confidence interval when analyzing all populations and populations from each region separately (Fig. 2-20). However, only the SNPs NAC_854 and NAC _962 (2.8% of all SNPs) remained significant after the FDR correction. These two SNPs are likely under positive selection. In the Arlequin analysis, five SNPs fell outside the 95% interval (Fig. 2-21). Even though they did not remain significant after FDR correction, three of them (CysPro_202, NAC_962 and 92_352) were also identified by LOSITAN (Table 2-10). Similar to the saplings, no significant outliers were detected by BayeScan for the adults.

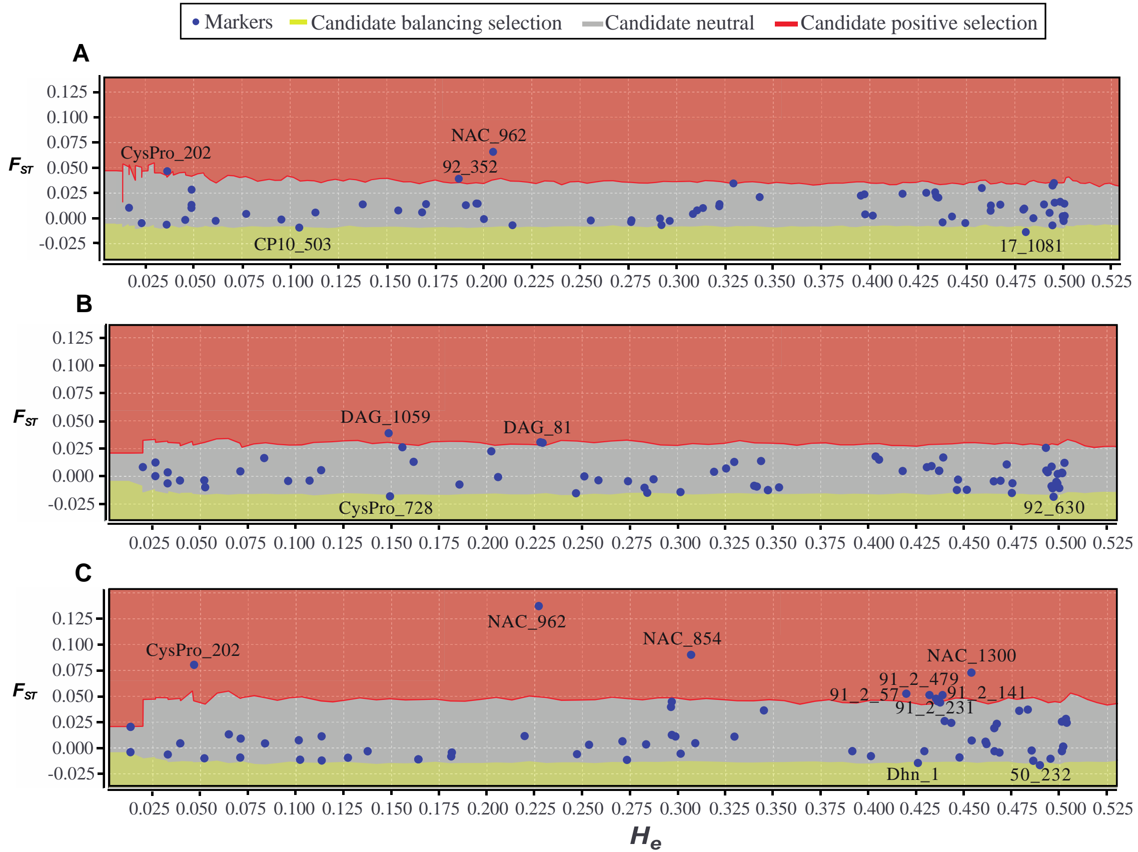


Fig. 2-20 Outlier SNPs detected in adults with LOSITAN under an island model. The plots show the distribution of the estimated F_{ST} values for each SNP along their mean heterozygosities (H_e) in **A,** all populations; **B,** populations from the Rhine valley; **C,** populations from the Rhone valley. The gray area represents the 95% confidence interval; SNPs falling outside that interval are labeled and could be potentially under selection.

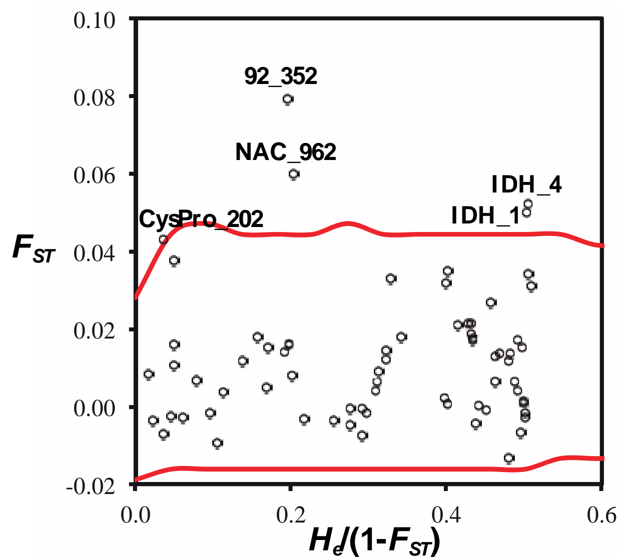


Fig. 2-21 ARLEQUIN results showing the distribution of the estimated F_{ST} values for each SNP along their mean standardized heterozygosities (H_e /(1-F_{ST})) in populations of adults. The loci falling outside the 95% interval were labeled and could be potentially under selection.

Among the detected outliers, the SNP 92_352 is likely a true one, because it was detected by both LOSITAN and Arlequin methods in the adults (Table 2-10). In addition, the SNPs CysPro_202 and NAC_962 were detected by LOSITAN and Arlequin not only in adults, but also by Arlequin in saplings, which makes them likely true outliers under selection (Table 2-10). The rest of the outliers could be false.

Table 2-10 List of the SNPs detected as outliers by one or two methods in saplings and adults.

Saplings			Adults		
LOSITAN	**ARLEQUIN**	**Selection**	**LOSITAN**	**ARLEQUIN**	**Selection**
ALDH_4*		Balancing			
	ERD	Positive			
				IDH_1	Positive
				IDH_4	Positive
			Dhn_1		Balancing
			CP10_503		Balancing
	CysPro_202	Positive	**CysPro_202**	**CysPro_202**	Positive
			CysPro_728		Balancing
			DAG_1059		Positive
			NAC_854*		Positive
	NAC_962	Positive	**NAC_962***	**NAC_962**	Positive
			NAC_1300		Positive
			17_1081		Balancing
			50_232		Balancing
			91_2_57		Positive
			91_2_141		Positive
			91_2_231		Positive
			91_2_479		Positive
			92_352	**92_352**	Positive
			92_630		Balancing

* The loci that remained significant after multiple testing correction. The loci highlighted by the bold font were detected by both methods. BayeScan did not detect any outliers.

In the SNP CysPro_202, representing a synonymous substitution, allele A has a frequency close to zero in almost all populations, but its frequency is higher in Chamoson in both adult and sapling populations (Figs. 2-22 A and B). A similar and even stronger tendency is also observed in the synonymous SNP NAC_962, where allele A has a higher frequency in Chamoson (Figs. 2-22 C and D) in both adults and saplings. In contrast, 92_352, a non-synonymous SNP, showed a difference in allele frequencies between the two valleys: in adult populations, allele A has an average frequency of 0.83 in the Rhine valley, and an

average frequency of 0.96 in the Rhone valley (Fig. 2-22E; Fisher exact two-tailed $P < 0.001$).

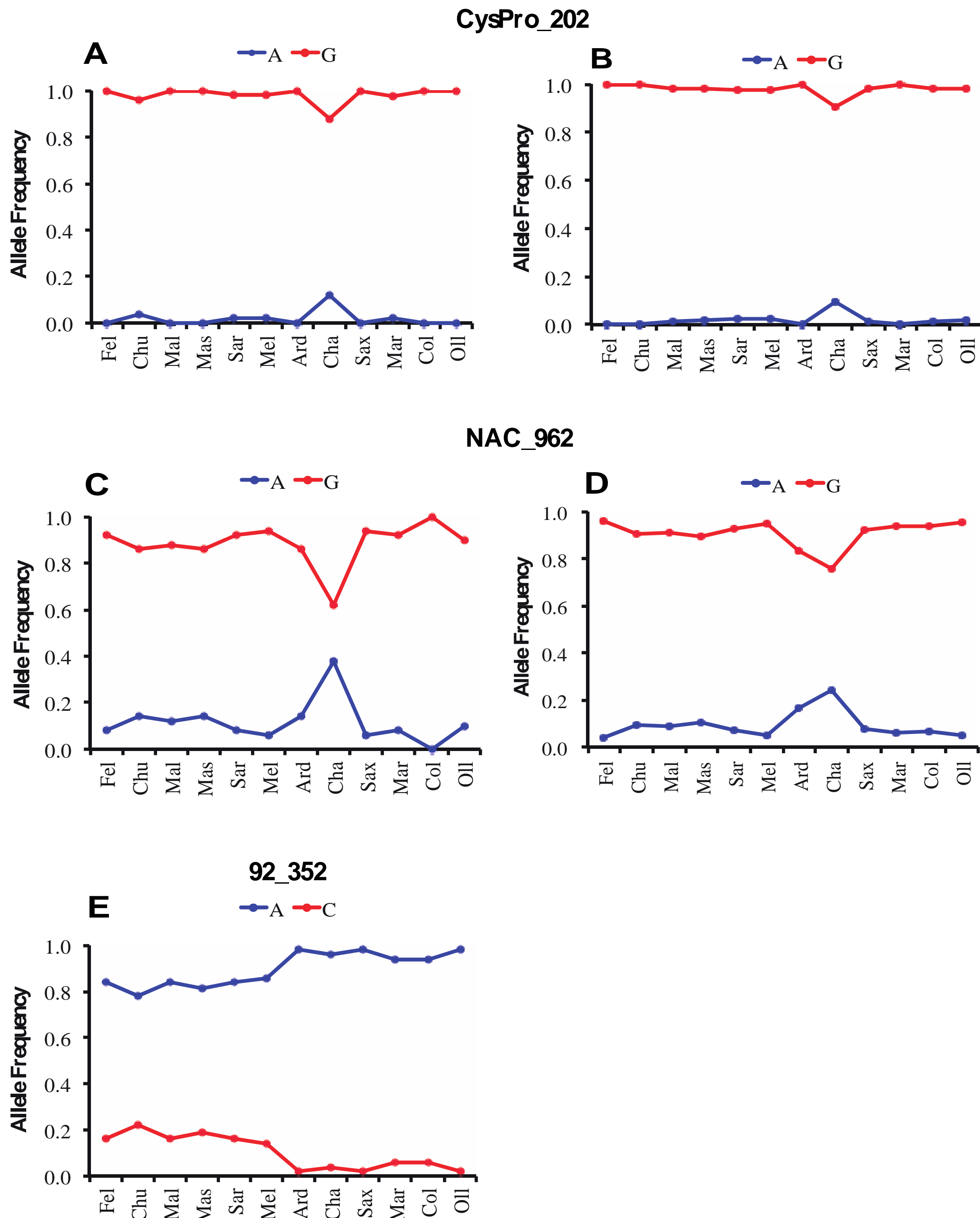


Fig. 2-22 The allele frequency distribution for the outlier SNPs under positive selection. **A, C,** and **E,** adults; **B** and **D,** saplings.

Discussion

Genetic variability and linkage disequilibrium (LD)

Results demonstrate that the studied populations of *F. sylvatica* have high genetic variability found in both SSR and SNP markers (Tables 2-4 and 2-8). No significant differences in genetic variability between saplings and adults were found, suggesting that the saplings represent the genetic variability of the adult populations. Similar levels of genetic variability have been found in other studies based on similar sets of SSR loci (Seifert 2012; Bontemps *et al.* 2013; Müller 2014; Rajendra *et al.* 2014), and slightly lower when compared to the studies based on other SSR loci (Buiteveld *et al.* 2007; Kraj & Sztorc 2009; Chybicki *et al.* 2009; Bilela *et al.* 2012). Among SSRs, EST-SSRs presented lower variability than genomic SSRs (Table 2-3). Similar results have been reported in other studies (Seifert 2012; Müller 2014) and can be attributed to the location of EST-SSRs in coding regions, making them more conserved and thus, less polymorphic (Varshney *et al.* 2005; Ellis & Burke 2007). SNP markers also revealed high genetic variability for the studied populations of European beech, comparable to the genetic variability found with other studies (Seifert *et al.* 2012; Müller *et al.* 2015b). It is known that this high genetic variability, characteristic of woody plants, is due to their large geographic ranges, long lives, outcrossing breeding systems and wide pollen dispersal (Hamrick *et al.* 1992).

Null alleles are alleles that fail to amplify due to mutations in the primer binding site, causing a bias in the estimation of allele frequencies and a reduction in observed heterozygosity (Ellis & Burke 2007). Additionally, they are more likely to occur when SSR loci are transferred from other species. Although 7 SSR loci used in this study were transferred from *F. crenata* and *Q. robur*, no loci showed evidence of null alleles, which is supported by the fixation indices (F_{IS}) close to zero (Table 2-3). Likewise, only few loci showed significant deviations from Hardy-Weinberg equilibrium, and this was due to an excess of heterozygous. These results confirm the observations from other studies indicating that the transferability of SSR loci among species of the genus *Fagus* is relatively high (Pastorelli *et al.* 2003; Lefèvre *et al.* 2012) and that transferability of EST-SSR can be successful even in species from different genus but the same family (Ellis & Burke 2007), as was the case for the EST-SSR transferred from *Q. robur*.

LD between SSR loci was found for 19.2% of all the possible pairs in the saplings. In contrast, 0.013% of all the possible pair combinations were found in LD for the adults, which is comparable to the low percentage found in other study (Lefèvre *et al.* 2012). The higher percentage of SSR loci in LD in the saplings could be an effect of relatedness, since groups of 2-4 saplings were collected underneath the same adult tree. In fact, those saplings had higher pairwise relatedness coefficient than saplings collected under different trees (see chapter 5). As for SNPs, 5.5% and 4.4% of all the possible pairs were found to be in LD in the saplings adults, respectively. These values are comparable to the percentage (5.01%) reported by (Pluess *et al.* 2016), and considerably lower than the percentages reported in other studies. For example, Müller *et al.* (2015a) observed LD for 18.45% of all possible SNP pairs. In general, low LD can be expected for a highly outcrossing, wind-pollinated tree species such as European beech (Jump *et al.* 2006; Aitken *et al.* 2008).

Genetic differentiation and population structure

The low G''_{ST} values and the inferred population structure demonstrated that there is weak population differentiation in the studied populations of *F. sylvatica* (Figs. 2-3, 2-4, 2-15 and 2-16). The AMOVA analysis also showed that only 1-2% of the genetic variability was among populations (Tables 2-5 and 2-9). Low genetic differentiation is also reflected in the low bootstrap values observed in the NJ trees based on the SSR or SNP markers (Figs. 2-5 and 2-17). In general, the clustering based on the SSR or SNP markers reflected neither the geographic distances between populations nor the environmental conditions in which they occur, although there was a tendency for populations from the same valley to cluster together. These findings are in consensus with other studies in beech that also reported low genetic differentiation in Germany (Sander *et al.* 2000; Rajendra *et al.* 2014; Müller *et al.* 2015a), Italy (Paffetti *et al.* 2012), France (Csilléry *et al.* 2014) and other parts of Europe (Buiteveld *et al.* 2007). High gene flow may explain the low differentiation, since *F. sylvatica* is an outcrossing wind-pollinated tree species with high rates of pollen flow among populations (Oddou-Muratorio *et al.* 2011; Piotti *et al.* 2012). In fact, beech pollen can travel for thousands of kilometers, from Germany and North Italy to Catalonia in Spain (Belmonte *et al.* 2008). This high pollen dispersal capability can explain the low genetic differentiation, even between populations from the two different valleys.

However, despite the low genetic differentiation in general, both sapling and adult xeric populations in the Rhone valley were more genetically differentiated. Among the saplings,

Martigny and Chamoson were the populations with the highest pairwise genetic differentiation based on SSRs (Fig. 2-2), and Chamoson based on SNP markers (Fig. 2-14). Likewise, STRUCTURE analysis with SSRs identified Chamoson as a genetically distinct population (Fig. 2-3). Interestingly, Chamoson was also the most differentiated population based on both SSR and SNP markers in the adults (Figs. 2-2 and 2-14).

Outlier analysis

The three different outlier detection methods detected different sets of outlier loci (Tables 2-6 and 2-10), although there were a few loci that were detected by more than one method. Discrepancies between different outlier detection methods are common and have been reported also in other studies (e.g., Russello *et al.* 2012; Tsumura *et al.* 2014; Konijnendijk *et al.* 2015). It can be attributed, on the one hand, to the different demographic assumptions underlying each method, and, on the other hand, to the different rates of type I (false positives) and type II (false negatives) errors (Narum & Hess 2011).

Additionally, different outlier loci were detected between saplings and adults (Tables 2-6 and 2-10). Not only can the environment exert different selection pressures at different life stages (Petit & Hampe 2006), but also different sets of genes are involved in the same trait at different stages (Prunier *et al.* 2013). Therefore, SNPs under selection are likely to differ between different ages. Besides, due to high competition and mortality, only a small fraction of seeds survive until the adult stage (Petit & Hampe 2006), which means that adult trees have passed different selection pressures through their life, and this could be evident in the higher number of outlier loci observed.

Interestingly, BayeScan detected as many outliers under balancing selection as almost a total number of SSR loci studied (92.3%) (Table 2-6), while no SNPs were identified as outliers (Table 2-10). Foll & Gaggiotti (2008) assumed that SSRs are a better choice to detect balancing selection, due to their higher polymorphism. However, SSRs have a high mutation rate and often underestimate F_{ST}, which could be inaccurately interpreted as a signature of balancing selection (Excoffier *et al.* 2009). Furthermore, the most of type I errors occur for balancing selection (Narum & Hess 2011). Thus, the results obtained with BayeScan must be interpreted carefully, since they are very likely to be false positives.

We used 10 SSRs that are located in non-coding regions (genomic SSRs) and 3 SSRs located in coding regions (EST-SSRs). Those located in non-coding regions are often considered selectively neutral (Holderegger *et al.* 2006). Nevertheless, they could be in LD

with closely linked locus or loci under selection. In this study, 3 SSR loci (23%) were identified as outliers by at least two methods in the adults (Table 2-6). Two of them, FS3-04 and csolfagus_06, are genomic SSRs, and thus, were very likely linked with a locus under selection. Particularly, FS3-04 has also shown evidence of positive selection in populations of beech occurring in different conditions of humidity and temperature (Bilela *et al.* 2012), strongly suggesting that this locus could be linked with a locus associated with local adaptation to these conditions. In this study, one of the alleles of FS3-04 (Fig. 2-11A) had a high frequency in some of the populations with low and intermediate amount of precipitation. Since this tendency was not necessarily observed in geographically close populations, this suggests that similar environmental conditions, including precipitation, may be responsible for this pattern. Unlike csolfagus_06 and FS3-04, FIR065 is an EST-SSR, but according to BLAST search the function of the sequence is unknown.

Unlike SSRs, SNPs in candidate genes are considered the best choice to detect signatures of selection, since they are directly located in coding regions. Different results were obtained with the three outlier methods, and BayeScan did not detect any outliers (Table 2-10); BayeScan is considered more conservative in identifying outlier SNPs than other methods (Narum & Hess 2011). In total, 3 SNPs (4.3%) were detected as outliers under positive selection by at least two methods in the adults; two of them were also detected in the saplings. The small proportion of outlier loci detected is in line with other studies carried out in forest trees such as *Cryptomeria japonica* (Tsumura *et al.* 2014) and *Quercus petraea* (Alberto *et al.* 2013). However, outlier methods have a limited sensitivity to identify markers under weak selection (Narum & Hess 2011). Furthermore, if there are subtle changes in allele frequencies, such as in the case of polygenic traits, in which adaptation involves subtle changes in allele frequencies at the loci controlling the polygenic trait (Stephan 2016), or when there is a high gene flow counteracting selection signatures (Rellstab *et al.* 2015), it can be difficult to detect outliers.

Two of the outlier SNPs detected in this study were synonymous: CysPro_202 and NAC_962. Nevertheless, these loci were likely under positive selection. Positive selection occurs when an allele is beneficial, and its frequency increases in a population (Nielsen 2005; Vitti *et al.* 2013). In both SNPs, allele A has a higher frequency in Chamoson while allele G is almost fixed (with a frequency close to 1) in the rest of populations (Figs. 2-22 A, B, C, and D). This may indicate positive selection in favor of A alleles in Chamoson. However, these two SNPs did not show significant association with the environmental variables (see chapter 4). Additionally, this population was also found to be genetically

distinct based on SSRs and SNPs; thus, it cannot be ruled out that factors other than selection could be could be responsible for the differentiation of this population at these two SNPs.

In contrast, 92_352, a non-synonymous SNP, showed a difference in allele frequencies between the two valleys in the adults: allele A has an average frequency of 0.83 in the Rhine valley, and an average frequency of 0.96 in the Rhone valley (Fig. 2-22E). This SNP also showed significant association with environmental variables (see chapter 4); however, populations from the same valley had similar allele frequencies regardless the contrasting environmental conditions where they occur. Since the Rhine valley is at the east and the Rhone valley at the west of Switzerland, distinct colonization sources could explain the differences in the two valleys observed at SNP 92_352. Beech from the Slovenian and eastern Alps are considered as the main source areas for the colonization of central and northern Europe during the postglacial period (Magri *et al.* 2006), while beech from the western Alps might be responsible for the colonization of southern France up the Rhone (Delhon & Thiébault 2005; Magri 2008). If migration history explained differences in the two valleys, this should be observed also in other loci; however, this is not the case. This might suggest that particular environmental conditions in each valley that were not considered may be responsible for the differences at this locus.

Among the likely true SNP outliers observed in this study, none of them has shown evidence of selection in other studies. Interestingly, some of the SNPs that were considered could be false outliers have been found to be associated with important climate-related traits. That is the case of the SNPs CP10_503, CysPro_728, DAG_1059, NAC_854 and NAC_1300, which have been associated with bud burst in beech (Müller *et al.* 2015a). Similarly, SNP 92_630 has been found to be associated with climatic variables (Pluess *et al.* 2016), and SNPs 91_2_141, 91_2_231 and 91_2_479 have shown evidence of epistatic selection (Csilléry *et al.* 2014). Thus, further study of these SNPs involving other approaches such as polygenic and epistatic selection (Fu & Akey 2013) will help determine their participation in adaptation to different environmental conditions.

3. Phenotypic Association analysis

Introduction

Rising temperatures and changes in precipitation patterns as a consequence of climate change are predicted to affect forest ecosystems (Milad *et al.* 2011). For central Europe, climate change scenarios predict an increase in duration and intensity of droughts (Kovats *et al.* 2014). These factors likely will affect survival of forest trees (Allen *et al.* 2010), and thus, alter the structure, composition and distribution of forests (Crookston *et al.* 2010; Chmura *et al.* 2011).

European beech (*Fagus sylvatica*) is one of the most common species in Europe, and its distribution could be affected as a consequence of climate change (Kramer *et al.* 2010). Changes in the phenology of this species have been observed, registering an advancement of spring bud burst parallel with the global climatic trend (Badeck *et al.* 2004). This increases the probability of late frost damage, affecting especially the survival of seedlings and saplings (Packham *et al.* 2012). Besides, the predicted increment in frequency and duration of summer droughts will likely affect European beech distribution, since severe drought periods may be harmful for this species (Milad *et al.* 2011), and it could be overcompeted by more drought-tolerant trees such as *Quercus petraea* and *Pinus sylvestris* (Geßler *et al.* 2007; Friedrichs *et al.* 2009).

Despite its susceptibility to drought, morphological and physiological data indicate that European beech populations from dry sites could be more drought tolerant than populations from wet sites (Peuke *et al.* 2002; Rose *et al.* 2009; Arend *et al.* 2016b). Likewise, genetic differences between beech populations growing in environments with different water availability have been found (Pluess & Weber 2012). Recently, associations between SNPs in candidate genes and important climate related traits, such as bud burst, have been found (Müller *et al.* 2015a), and associations between SNPs and different environmental variables such as elevation (Csilléry *et al.* 2014), temperature, precipitation and drought, have been detected (Pluess *et al.* 2016) demonstrating standing genetic variation very much needed for local adaptation.

Given that precipitation gradients may reflect differences in water availability and thus, promote local adaptation to drought, the aim of this study is to assess the response of saplings from beech populations occurring along steep precipitation gradients to simulated summer drought conditions on two different types of soil. The phenotypic traits bud burst, chlorophyll fluorescence and stem growth were assessed, and the association between traits and SNPs in

genes that supposedly may control these adaptive traits (so called candidate genes) was tested.

Materials and Methods

Plant material

Twelve European beech populations growing along precipitation gradients in two inner alpine valleys in Switzerland were selected (Table 3-1). In spring 2011, 16-31 adult trees per population about 50 m apart from each other were selected, and 2-4 saplings underneath them with a size of ~20 cm were excavated. In total, 755 saplings were collected and transplanted to the model ecosystem facility MODOEK of the Swiss Federal Institute for Forest, Snow and Landscape Research WSL.

Table 3-1 Environmental characteristics of the selected populations. Climate data were taken from nearby METEO SWISS stations (distance ≤ 10km) for the 1981-2010 period.

Valley	Population	*N*	Position	Elevation, m.a.s.l	Mean annual	
					temperature, °C	precipitation, mm
Rhine	Felsberg	62	46°51'N, 9°28'E	650-800	10.0	849
	Chur	63	46°52'N, 9°32'E	700-800	10.0	849
	Malans	64	46°59'N, 9°34'E	600-700	10.1	1114
	Mastrils	62	46°58'N, 9°32'E	550-650	10.1	1114
	Sargans	63	47°3'N, 9°26'E	650-750	10.1	1334
	Mels	60	47°3'N, 9°24'E	650-750	10.1	1334
Rhone	Ardon	63	46°13'N, 7°14'E	750-850	10.1	603
	Chamoson	64	46°12'N, 7°12'E	750-850	10.1	603
	Saxon	64	46°8'N, 7°11'E	700-800	10.1	603
	Martigny	64	46°6'N, 7°6'E	500-700	10.1	855
	Collombey	63	46°16'N, 6°56'E	550-650	9.8	1012
	Ollon	63	46°18'N, 6°59'E	600-700	9.8	1012

N - Number of individuals sampled

Experimental design

The drought experiment described in the following section was carried out by Dr. Matthias Arend, from the Swiss Federal Institute for Forest, Snow and Landscape Research WSL. The MODOEK facility of the WSL consists of 16 chambers equipped with a sliding roof and an automated irrigation system for controlling water supply. Each chamber is split below ground in two lysimeters containing acidic or calcareous forest soil with a pH of 4.0 and 6.9, respectively, and comparable texture. In each chamber, two saplings from each population were transplanted on each type of soil. Saplings were acclimatized during two growing

seasons (2011 and 2012) to overcome the transplantation shock and to regenerate their root system. The sliding roofs were closed from May to October to exclude natural precipitation, and the saplings were irrigated every second or third day with deionized water enriched with nutrients simulating the average composition of ambient rainfall (Kuster *et al.* 2013). During hot periods, the intensity and frequency of irrigation was increased to hold soil moisture and counterbalance high rates of evapotranspiration.

In 2013 and 2014, the drought experiment was carried out. Drought conditions were imposed in half of the chambers by omitting water irrigation from May to August. During hot days, short irrigation pulses were applied to avoid too intense soil drying and, irreversible damage of the saplings due to water loss by high evapotranspiration. The other half of the chambers were used as control, maintaining irrigation as described above. A summary of the number of saplings under the different conditions of soil and treatment is found in Table 3-2.

Table 3-2 Summary of the number of saplings under different conditions

	Control	Drought	Total
Acidic soil	188	193	381
Calcareous soil	187	187	374
Total	375	380	755

Phenotypic data

The following phenotypic traits measured in the saplings were evaluated and provided by Dr. Matthias Arend from the Swiss Federal Institute for Forest, Snow and Landscape Research WSL.

Spring bud burst was assessed daily in 2012, 2013 and 2014. A sapling was considered to be bursting when the first bud had broken out and the green leaf was visible. Date of bud burst was recorded and transformed into day of year. Bud burst was assessed before the onset of the drought experiment on May (Arend *et al.* 2016a).

To assess the physiological responses of the saplings to the drought experiment, chlorophyll fluorescence was evaluated in 2014 on the saplings growing on acidic soil. Fast fluorescence kinetics was analyzed once per sapling between 11:00 and 12:00 on dark-adapted leaves using a portable plant efficiency analyzer (Pocket PEA, Hansatech Instruments Ltd., Norfolk, UK). After a saturating light pulse of 3500 µmol quanta m^2/s of red light (650 nm), increase in fluorescence was registered. Based on fluorescence kinetics,

three parameters were calculated: The maximum quantum efficiency of PSII (F_V/F_M), the performance index of PSII on absorption basis (PI_{abs}), and the total performance index of PSII (PI_{tot}) (Arend *et al.* 2016b).

Additionally, radial stem growth (SG) 0.5 cm above ground was evaluated in all saplings as a morphological response to the experiment. For that purpose, diameter of the stems was measured in March 2013, October 2013 and September 2014, and the increment in diameter was calculated as the difference between March-October 2013 (SG 2013), October 2013-September 2014 (SG 2014) and March 2013-Septermber 2014 (SG 2013-2014).

Statistical analysis of phenotypic data

Since non-normal distribution of the phenotypic data was revealed by the Kolmogorov-Smirnov test, the non-parametric Kruskall-Wallis test was used to test for significant differences on phenotypic traits between drought and control treatments and acidic and calcareous soil. Furthermore, statistical differences between the most mesic (Sargans and Mels) and most xeric populations (Ardon, Chamoson and Saxon) under drought conditions were also tested. Statistical analyses were performed with R 3.3.1 (R Core Team 2016).

SNPs and candidate genes

Some of the SNPs in candidate genes involved in phenology and drought stress tolerance from previously published studies for *F. sylvatica* (Seifert *et al.* 2012; Lalagüe *et al.* 2014; Müller *et al.* 2015b) were selected as follows: for each gene, linkage disequilibrium (LD) blocks were identified by using the software htSNPer 1.0 (Ding *et al.* 2005). Then, a subset of SNPs representing the majority of haplotypes (haplotype tag SNPs) was selected for further genotyping. In addition, SNPs showing signatures of natural selection in the previous studies (Csilléry *et al.* 2014; Müller *et al.* 2015a) were also selected, for a total of 24 genes and 76 SNPs (21 non-synonymous, 27 synonymous and 28 in non-coding regions) selected for genotyping (Table 3-3). Nucleotide sequences neighboring selected SNPs were sent to LGC Genomics Ltd. for primer design and SNP genotyping using the PCR-based KASP™ genotyping assay (Hoddesdon, UK).

Table 3-3 Candidate genes and characteristics of the selected SNPs

Gene	SNP name	Type	Reference
Aldehyde dehydrogenase	ALDH_1	Non-coding	Seifert *et al.* 2012
	ALDH_2	Non-Synonymous	
	ALDH_3	Non-Synonymous	
	ALDH_4	Synonymous	
Isocitrate dehydrogenase	IDH_1	Synonymous	
	IDH_3	Non-coding	
	IDH_4	Synonymous	
Ascorbate peroxidase	APX1_1	Synonymous	
	APX1_2	Non-coding	
	APX4_1	Non-coding	
	APX4_2	Non-Synonymous	
Early responsive to dehydration	ERD	Non-coding	
Dehydrin	Dhn_1	Non-Synonymous	
	Dhn_2	Non-Synonymous	
Glutathione peroxidase	GPX	Non-Synonymous	
Phytochrome B	PhyB	Synonymous	
Cysteine proteinase	CysPro_118	Synonymous	Müller et al. 2015
	CysPro_202	Synonymous	
	CysPro_728	Non-coding	
	CysPro_783	Non-coding	
Chloroplast Chaperonin like	CP10_65	Synonymous	
	CP10_67	Non-Synonymous	
	CP10_377	Non-coding	
	CP10_442	Non-coding	
	CP10_503	Synonymous	
	CP10_749	Synonymous	
	CP10_1317	Non-coding	
	CP10_1428	Non-Synonymous	
Dof zinc finger protein	DAG_81	Non-coding	
	DAG_289	Non-coding	
	DAG_1059	Synonymous	
Histone 3	His3C1_292	Non-coding	
	His3C2_104	Synonymous	
	His3C2_186	Non-coding	
	His3C2_260	Synonymous	
NAC transcription factor	NAC_854	Non-Synonymous	
	NAC_962	Synonymous	
	NAC_1300	Non-coding	
Protein phosphatase 2C	PP2C_315	Non-Synonymous	Müller et al. 2015
	PP2C_391	Synonymous	
	PP2C_791	Non-Synonymous	
	PP2C_941	Non-coding	
	PP2C_1200	Synonymous	
Xyloglucan endotransglucosylase/hydrolase 23	7_258	Non-coding	Lalagüe et al. 2014
	7_520	Non-coding	
Short chain alcohol dehydrogenase	17_880	Non-coding	

Gene	SNP name	Type	Reference
	17_1081	Non-coding	
Potassium transporter 2	39_256	Synonymous	
	39_282	Non-Synonymous	
CRT/DRE binding factor	50_39	Non-Synonymous	
	50_232	Synonymous	
	50_320	Non-coding	
s-adenosyl-l-homocysteine hydrolase	52_1_235	Non-Synonymous	
	52_1_249	Non-Synonymous	
	52_1_368	Synonymous	
Glyceraldehyde 3-phosphate dehydrogenase	68_277	Non-Synonymous	
	68_313	Non-coding	
Light-harvesting complex II protein	88_1_450	Non-coding	
	88_1_727	Synonymous	
	88_1_803	Non-Synonymous	
Catalase	91_2_57	Synonymous	
	91_2_141	Synonymous	
	91_2_231	Synonymous	
	91_2_448	Non-coding	
	91_2_479	Non-coding	
	91_2_504	Non-coding	
1-aminocyclopropane-1-carboxylate oxidase	92_166	Non-coding	
	92_352	Non-Synonymous	
	92_630	Non-coding	
Cytosolic class I small heat-shock protein	110_1_111	Non-Synonymous	
	110_1_293	Synonymous	
	110_1_423	Non-Synonymous	
	110_1_450	Non-Synonymous	
Pectin methylesterase	154_2_137	Synonymous	
	154_2_371	Synonymous	
	154_2_617	Synonymous	

Association analysis

The software TASSEL 5.0 (Bradbury *et al.* 2007) was used to test for association between SNPs and the following phenotypic traits: day of bud burst, F_V/F_M, PI_{abs}, PI_{tot} and SG. Association analyses were done by grouping individuals according to the experimental conditions: treatment, soil and treatment/soil. In addition, an analysis by grouping all saplings regardless of experimental conditions was also done; in that case, to account for the variability due to the different experimental conditions, normalization was applied by dividing the data by the mean of each experimental condition. Association analyses were performed using the general linear model (GLM) considering the population structure (*Q*) obtained from microsatellite data as a confounding factor using the software STRUCTURE

2.3.4 (Pritchard *et al.* 2000). Additionally, the mixed linear model (MLM) considering both *Q* and kinship (*K*) as confounding factors was also used. Pairwise *K*-matrix was estimated using relatedness coefficients according to Queller & Goodnight (1989) using the software GenAlEx 6.5 (Peakall & Smouse 2006, 2012), and negative values were set as zero. Correction for multiple testing was done by two different methods: Bonferroni correction ($P \leq 0.00143$) and adjustment of *P* values for a false discovery rate FDR < 0.1 using the Benjamini & Hochberg (1995) method implemented in the R function "p.adjust" (R Core Team 2016).

Results

Phenotypic data

Day of Bud Burst

Timing of bud burst was highly variable among individuals from the same population and among populations. Significant differences were found among populations in 2012, 2013 and 2014, being 2014 the year with the earliest bud burst for all populations (Fig. 3-1). Likewise, statistically significant differences were found between saplings growing on acidic and calcareous soil (Fig. 3-2), with a tendency for saplings on acidic soil to flush earlier in 2013 and 2014. No significant differences in day of bud burst were found between the most mesic (Sargans and Mels) and most xeric populations (Ardon, Chamoson and Saxon).

Chlorophyll Fluorescence

Physiological responses to the treatment measured by chlorophyll fluorescence were very variable among individuals, especially for the parameter PI_{abs}, which varied from 0.07 to 11.39 in control saplings, and from 0.17 to 7.21 in saplings under drought (Fig. 3-3B). In contrast, the parameter F_V/F_M varied from 0.48 to 0.85 and from 0.55 to 0.83 in saplings under control and drought treatment, respectively (Fig. 3-3A). No significant differences were found among populations for the different parameters (Fig. 3-3); however, differences were almost significant between mesic and xeric populations under drought treatment for the parameter PI_{tot}, with xeric populations performing better (P=0.06, Fig. 3-4). On the other hand, significant differences were found between control and drought treatments only with

the indices PI_{abs} and PI_{tot} (Fig. 3-5); on average, control plants performed 1.2 and 1.3 times better according to PI_{abs} and PI_{tot} indices, respectively.

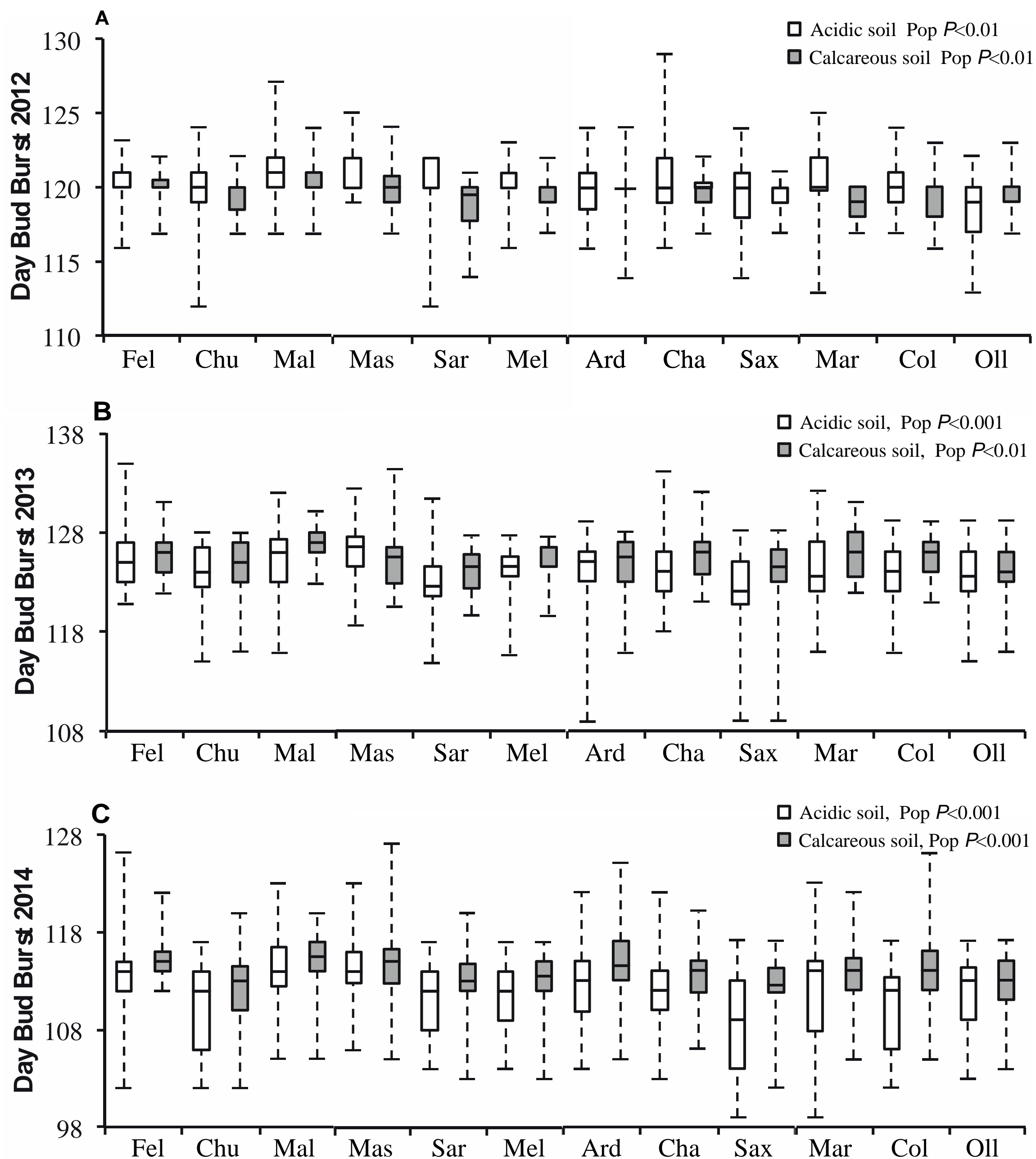


Fig. 3-1 Day of bud burst of the single populations growing on acidic and calcareous soil, in different years. **A**, 2012; **B**, 2013; **C**, 2014. Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values. Fel - Felsberg; Chu - Chur; Mal - Malans; Mas - Mastrils; Sar - Sargans; Mel - Mels; Ard - Ardon; Cha - Chamoson; Sax - Saxon; Mar - Martigny; Col - Collombey; Oll - Ollon.

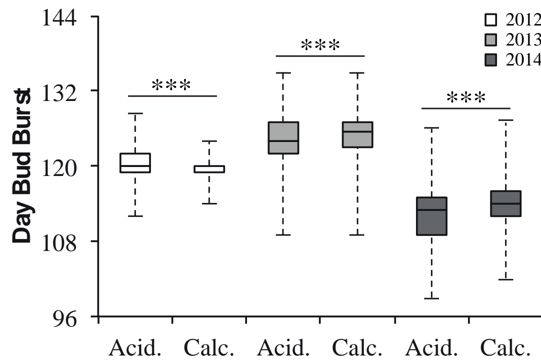


Fig. 3-2 Day of bud burst over all populations by type of soil in different years. *** $P < 0.001$. Acid. - acidic soil; Calc. - calcareous soil. Box plots include median and upper and lower quartiles; whiskers show minimum and maximum values.

Fig. 3-3 Physiological responses of the single populations to control and drought treatments measured by chlorophyll fluorescence parameters. **A**, F_V/F_M; **B**, PI_{abs}; **C**, PI_{tot}. Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values. Fel - Felsberg; Chu - Chur; Mal - Malans; Mas - Mastrils; Sar - Sargans; Mel - Mels; Ard - Ardon; Cha - Chamoson; Sax - Saxon; Mar - Martigny; Col - Collombey; Oll - Ollon.

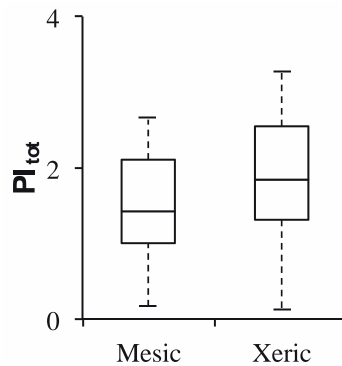


Fig. 3-4 Responses of mesic and xeric populations to the drought treatment, measured by the parameter PI_{tot}. Differences were almost significant (P=0.06). Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values.

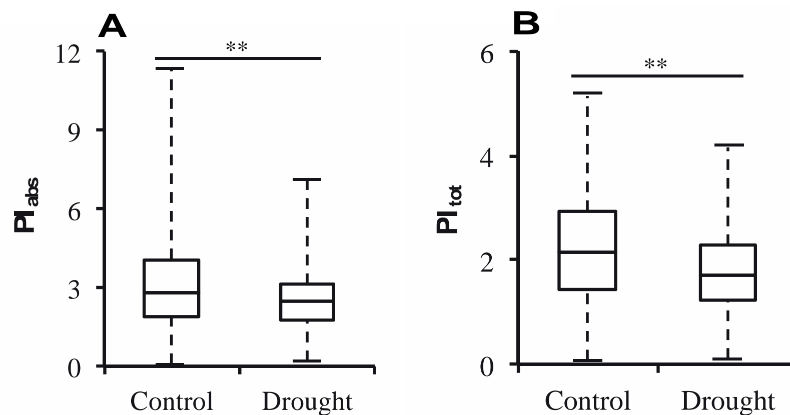


Fig. 3-5 Responses over all populations to control and drought treatments measured by **A**, PI_{abs} and **B,** PI_{tot}. **P<0.01. Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values.

Stem Growth

There was a high variability on stem growth as a response to the treatment and type of soil (Figs. 3-6 and 3-7). Differences in the response among populations were significant in control/calcareous, drought/acidic and drought/calcareous conditions (Figs. 3-6 and 3-7). Soil had a significant effect on plants under control and drought conditions in 2014 (Fig. 3-8B), and also on plants under drought conditions in the overall stem growth 2013-2014 (Fig. 3-8C); saplings on calcareous soil had the highest stem growth (Figs. 3-8B and 3-8C). Treatment had a significant effect on SG in 2013, 2014 and 2013-2014 (Fig. 3-8); control plants had higher stem growth. Mesic and xeric populations demonstrated significant differences in stem growth under drought conditions in both acidic and calcareous soil (Fig. 3-9). In all cases, xeric populations demonstrated higher values of stem growth.

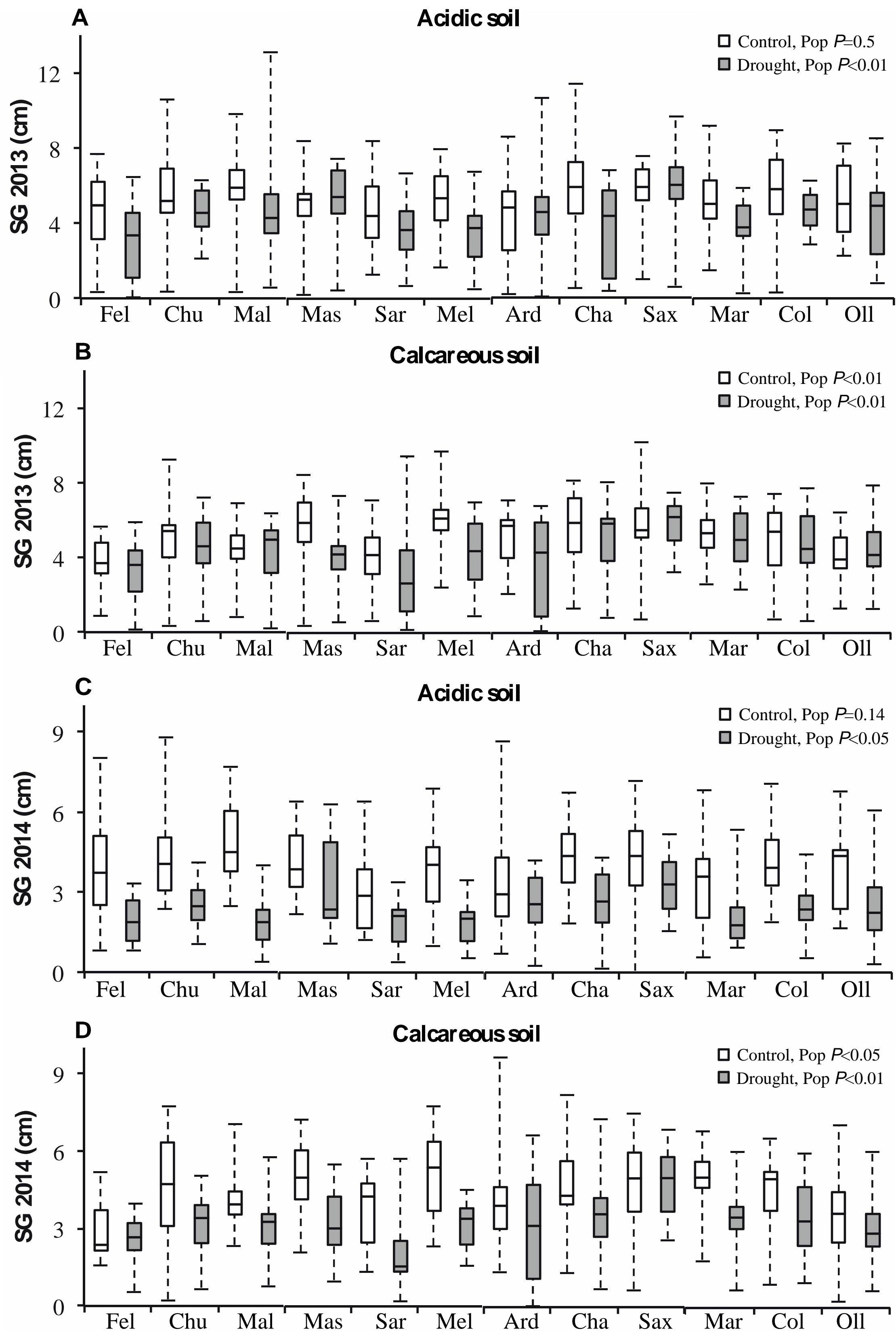


Fig. 3-6 Morphological responses of the single populations to treatment and type of soil, measured as stem growth (SG) in 2013 and 2014. Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values. Fel - Felsberg; Chu - Chur; Mal - Malans; Mas - Mastrils; Sar - Sargans; Mel - Mels; Ard - Ardon; Cha - Chamoson; Sax - Saxon; Mar - Martigny; Col - Collombey; Oll - Ollon.

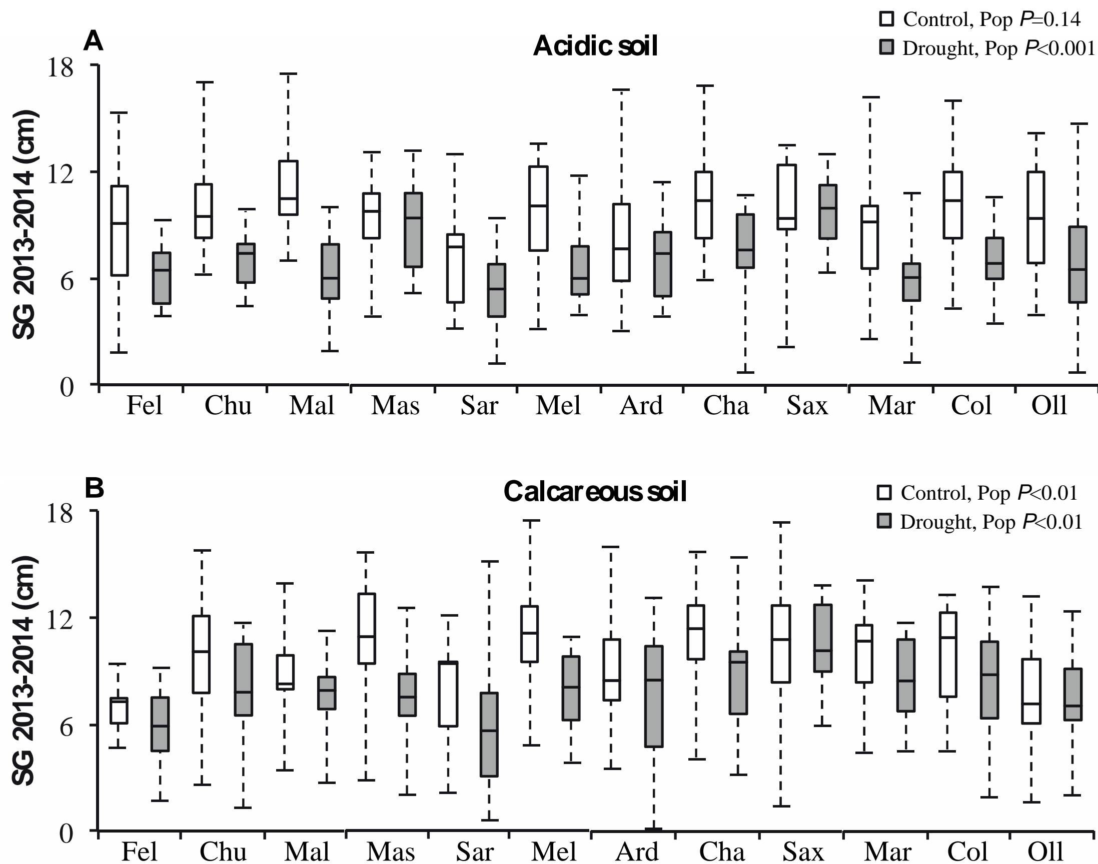


Fig. 3-7 Morphological responses of the single populations to treatment and type of soil, measured as overall stem growth (SG) 2013 - 2014. Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values. Fel - Felsberg; Chu - Chur; Mal - Malans; Mas - Mastrils; Sar - Sargans; Mel - Mels; Ard - Ardon; Cha - Chamoson; Sax - Saxon; Mar - Martigny; Col - Collombey; Oll - Ollon.

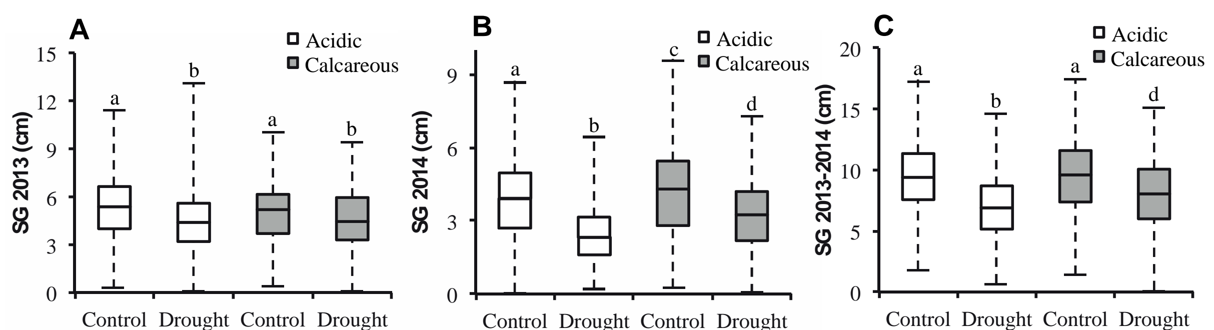


Fig. 3-8 Effect of treatment and type of soil on stem growth over all populations in **A**, 2013; **B**, 2014; **C**, 2013-2014. Different letters indicate significant differences, $P<0.05$. Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values.

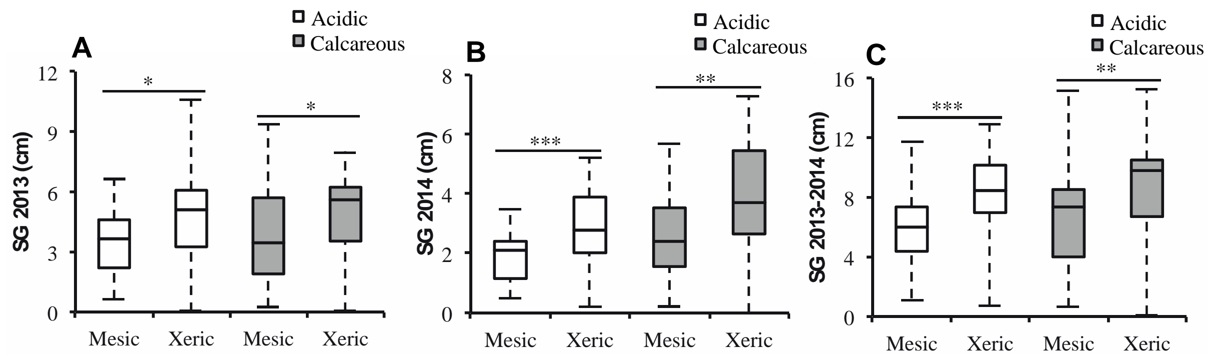


Fig. 3-9 Differences in the responses of mesic and xeric populations to the drought treatment measured by stem growth (SG) in **A**, 2013; **B**, 2014; **C**, 2013-2014. *P<0.05; **P<0.01; ***P<0.001. Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values.

Phenotypic Association analysis

From the 76 selected SNPs for genotyping, 6 were monomorphic (APX1_2, PhyB, 50_320, 52_1_249, 92_166, 110_1_111); thus, the remaining 70 were used for the association analysis. Comparison between the GLM and MLM methods revealed that they yielded similar results and had very similar performance, as shown by the close distribution between the observed and expected *P* values in the quantile-quantile plots (Appendix 3-1). Therefore, only the results corresponding to the MLM model are presented.

In total, 5 out of 70 SNPs analyzed showed significant association with the phenotypic traits PI_{abs}, PI_{tot} and SG; and 2 SNPs showed close to significance association with F_V/F_M (Table 3-4). None of the SNPs showed significant association with bud burst. Three of the significant SNPs showed association with more than one trait (50_39, IDH_1 and IDH_4). Only one of the SNPs showing association is located in non-coding regions; the rest of the SNPs are located in coding regions, two of them representing non-synonymous substitutions (PP2C_315 and 50_39), and four of them synonymous substitutions (110_1_293, IDH_1, IDH_4 and 50_232). The genotypic variation explained by the SNPs was relatively high ($4.1 \leq R^2 \leq 13.4$) (Table 3-4).

Homozygous plants CC for the SNP PP2C_315 showed higher F_V/F_M values (on average, 0.05 higher) than heterozygous (Figs. 3-10A and 3-11A). On the other hand, homozygous TT for the SNP 7_520 showed slightly lower values, on average 0.003 less, than heterozygous (Fig. 3-10B).

Table 3-4 Results of the association analysis under a MLM model for the pooled individuals (All saplings) and individuals under Drought/Acidic soil, Drought/Calcareous soil, and Control/Acidic soil conditions. Only SNPs showing significant or close to significant association with a phenotypic trait are presented

Trait	Gene	SNP	SNP type	All saplings			Drought/Acidic			Drought/Calcareous			Control/Acidic		
				R^2	*P*	*P**	R^2	*P*	*P**	R^2	*P*	*P**	R^2	*P*	*P**
F_V/F_M	*Protein phosphatase 2C*	PP2C_315	Non-synonymous	4.1	0.002^{+}	0.118^{+}	5.1	0.002^{+}	0.149^{+}	ND	ND	ND	2.9	0.143	0.940
	Xyloglucan endotransglucosylase hydrolase 23	7_520	Non-coding	4.1	0.002^{+}	0.118^{+}	7.3	0.003	0.188	ND	ND	ND	0.0	0.890	1.000
PI_{abs}	*Cytosolic class I small heat-shock protein*	110_1_293	Synonymous	5.8	**0.000**	**0.025**	3.8	0.066	0.789	ND	ND	ND	13.4	**0.000**	**0.020**
PI_{tot}	*CTR/DRE binding factor*	50_39	Non-synonymous	2.4	0.031	0.780	10.6	**0.000**	0.110^{+}	ND	ND	ND	1.3	0.502	1.000
SG 2013	*CTR/DRE binding factor*	50_39	Non-synonymous	0.3	0.471	0.995	8.5	**0.001**	0.138^{+}	0.0	0.960	0.985	3.6	0.091	0.911
SG 2014	*Isocitrate dehydrogenase*	IDH_1	Synonymous	1.4	0.014	0.723	0.4	0.851	0.990	10.2	**0.000**	**0.069**	0.4	0.853	1.000
		IDH_4	Synonymous	1.5	0.011	0.723	1.3	0.488	0.990	10.8	**0.000**	**0.069**	1.6	0.400	1.000
SG 2013-2014	*Isocitrate dehydrogenase*	IDH_1	Synonymous	1.1	0.045	0.869	0.0	0.998	0.999	8.8	**0.001**	0.111^{+}	0.0	0.999	1.000
		IDH_4	Synonymous	1.1	0.036	0.869	1.0	0.586	0.990	9.7	**0.001**	**0.069**	1.4	0.458	1.000
	CTR/DRE binding factor	50_39	Non-synonymous	0.1	0.882	1.000	8.2	0.002^{+}	0.138^{+}	0.1	0.918	0.983	0.3	0.900	1.000
		50_232	Synonymous	0.0	0.975	1.000	8.4	**0.001**	0.138^{+}	0.2	0.856	0.983	0.8	0.696	1.000

R^2 - phenotypic variation explained by the SNP (%). *ND* - no determined. *P** - adjusted *P*. In bold, significantly associated after applying Bonferroni correction ($P \leq 0.0014$) or after adjusting *P* values for a FDR<0.1. $^{+}$ - close to significance

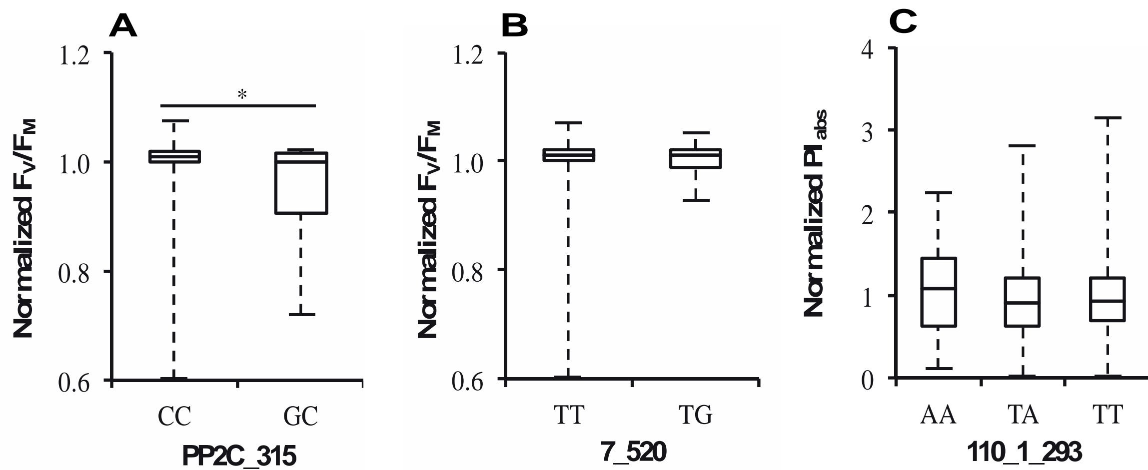


Fig. 3-10 SNPs showing association in the analysis using all saplings, and their effects on normalized chlorophyll fluorescence parameters. **P*<0.05. Homozygous GG in A and B are not shown due to low sample size (N=1). Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values.

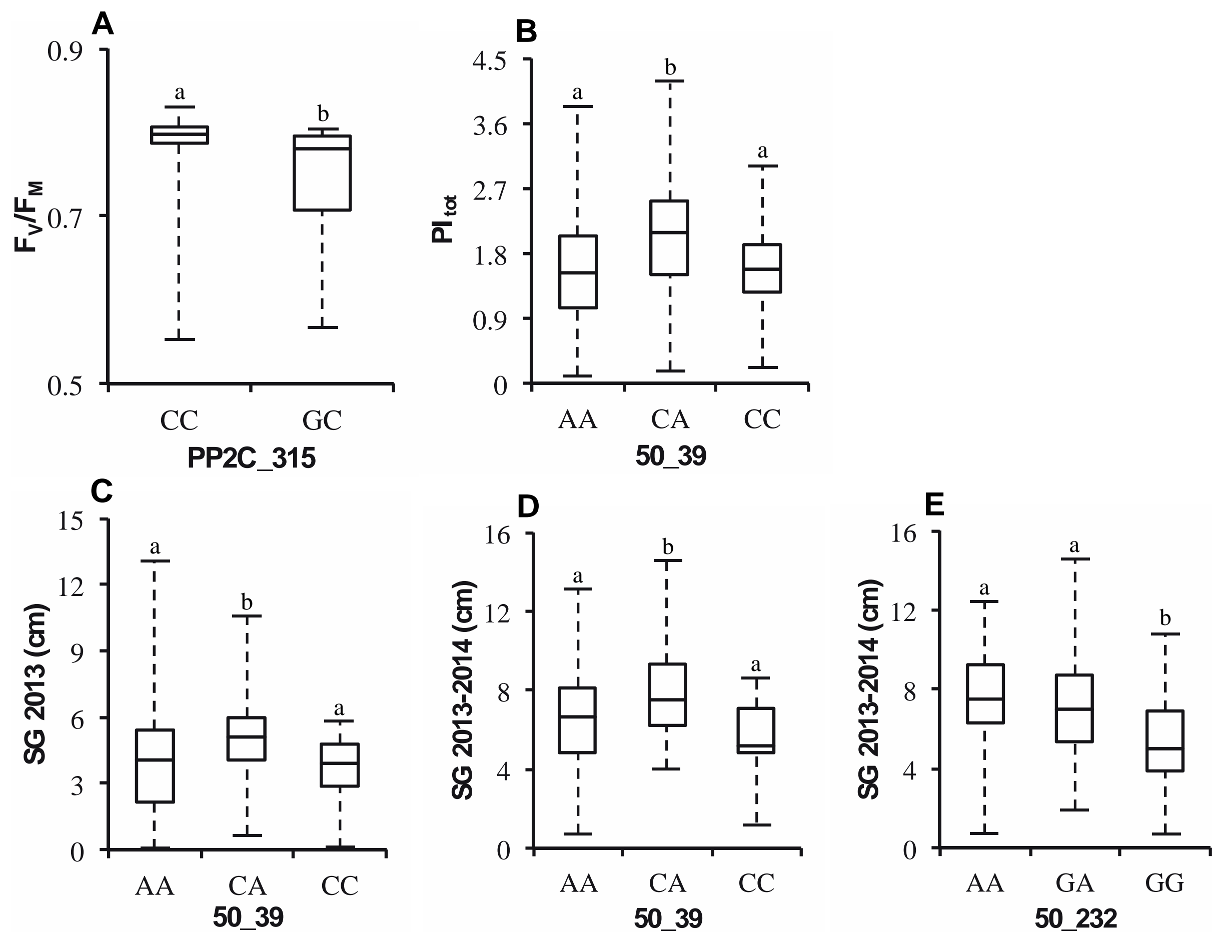


Fig. 3-11 SNPs showing association in saplings under drought/acidic soil conditions, and their effects on the phenotypic traits. Different letters indicate significant differences, *P*<0.05. Homozygous GG in A are not shown due to low sample size (N=1). Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values.

Homozygous individuals AA for SNP 110_1_293 showed higher normalized PI_{abs}: on average, 0.12 higher than heterozygous and 0.06 higher than homozygous TT (Fig. 3-10C). In contrast, in saplings on control/acidic soil conditions, homozygous TT showed higher PI_{abs}: on average 0.17 higher than heterozygous and 0.26 higher than homozygous AA (Fig. 3-12).

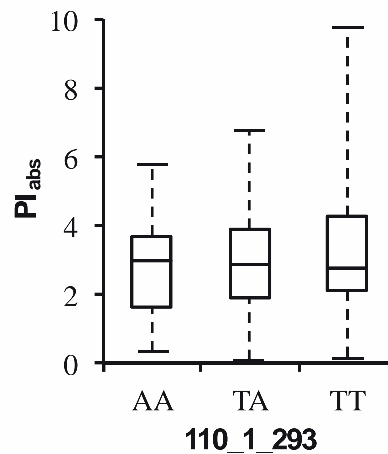


Fig. 3-12 SNP showing significant association in saplings under control/acidic soil conditions, and its effect on PIabs. Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values.

Heterozygous saplings for the SNP 50_39 showed higher values for the parameter PI_{tot} (0.48 higher than AA and 0.53 higher than CC) (Fig. 3-11B); also in SG 2013 (1.23 higher than AA and 1.64 higher than CC) (Fig. 3-11C); and in SG 2013-2014 (1.26 higher than AA and 2.05 higher than CC) (Fig. 3-11D). Homozygous saplings AA for SNP 50_232 showed higher SG 2013-2014, on average 0.22 higher than heterozygous and 1.94 higher than homozygous GG. However, the differences between AA and GA were not significant (Fig. 3-11E).

Regarding SNP IDH_1, heterozygous individuals had higher stem growth than homozygous CC (on average 0.09 higher in SG 2014 and 0.3 higher in SG 2013-2014), and higher than homozygous TT (on average 1.07 higher in SG 2014 and 1.94 in SG 2013-2014), the differences between CC and TC were not significant (Figs. 3-13 A and C). Similarly, within the same gene, heterozygous saplings for the SNP IDH_4 had higher stem growth, on average 1.04 and 1.88 higher than homozygous AA in SG 2014 and SG 2013-2014, respectively; and on average 0.09 and 0.32 higher than GG in SG 2014 and SG 2013-2014,

respectively. However, differences between GA and GG genotypes were not significant (Figs. 3-13 B and D).

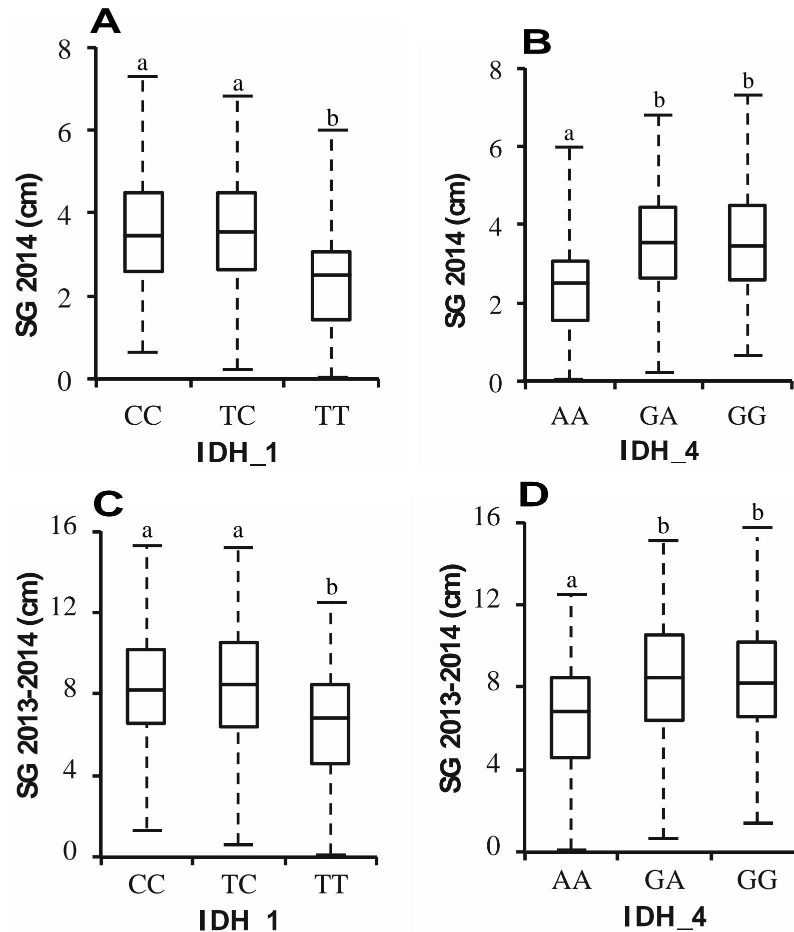


Fig. 3-13 SNPs showing significant association in saplings under drought/calcareous soil conditions, and their effects on stem growth. Different letters indicate significant differences, $P<0.05$. Box plots include median, upper and lower quartiles; whiskers show minimum and maximum values.

Discussion

European beech is one of the most important forest tree species in Europe, and its adaptive potential to climate change is of great interest. Important climate related-traits such as bud burst and drought tolerance are regulated not only by environmental signals but also by genetic factors, whose investigation is key for a better understanding of the genetic basis of important traits.

Day of Bud burst

Bud burst is an important phenological trait regulated by the interaction among chilling, forcing temperatures and photoperiod (Caffarra & Donnelly 2011; Basler & Körner 2014), and is one of the traits that has been affected the most by global warming (Schröder *et al.* 2006). In this study, timing of bud burst was highly variable among populations in all the three years in which it was assessed (Fig. 3-1). It is known that while western and northern populations show a clear trend towards late flushing, populations of beech in Central Europe show variability in the time of flushing (Jazbec *et al.* 2007). Furthermore, a clear earlier onset of bud burst was observed in 2014 (Fig. 3-2), which could be explained by the higher spring temperature during that year, a trend that has been already observed in beech with the rising temperatures (Badeck *et al.* 2004; Fu *et al.* 2012). Additionally, bud burst was also earlier in the saplings growing in acidic soil. Soil characteristics also influence timing of bud burst (Arend *et al.* 2016a). For example, high soil moisture favors early bud burst in birch, whereas high pH seems to delay it (Wielgolaski 2001).

Chlorophyll fluorescence

Chlorophyll fluorescence has been usually used to evaluate the state of photosystem II, and thus, the overall rate of photosynthesis, giving insights into plant responses to different stresses (Maxwell & Johnson 2000). Among the different chlorophyll fluorescence based parameters calculated to evaluate the performance of PSII, F_V/F_M was the least variable, and no significant differences among populations and control vs. drought treated saplings were found with this parameter (Fig. 3-3). This is in line with other studies reporting low variability and responsiveness of F_V/F_M to drought stress (Gallé & Feller 2007; Robson *et al.* 2009; Arend *et al.* 2016b; Cocozza *et al.* 2016).

In contrast, PI_{abs} and PI_{tot} were better indicators of drought stress sensitivity. Saplings under drought stress had lower values of PI_{abs} and PI_{tot}, indicating a negative effect on photosynthesis caused by water shortage (Fig. 3-5). Other studies have also shown a reduction in photosynthesis in beech saplings exposed to drought, due to reduced stomatal conductance and consequently reduced assimilation rate of CO_2 (Gallé & Feller 2007; Priwitzer *et al.* 2014). Additionally, the index PI_{tot} showed that saplings from xeric population were less affected by the drought treatment than saplings from mesic populations (Fig. 3-4), and there is evidence indicating that xeric populations recover faster after drought

(Arend *et al.* 2016b). By measuring parameters such as carbon isotropic composition, transpiration rate and water potential in leaves, shoots and roots, other studies have also demonstrated that beech trees from dry habitats are less affected by drought (Peuke *et al.* 2002; Fotelli *et al.* 2009).

Stem Growth

Stem growth is an important phenotypic trait for the evaluation of sensitivity to environmental stresses such as drought. In this study, high variability in SG was found among populations under the same treatment and soil conditions (Figs. 3-6 and 3-7). Large phenotypic variation has been reported for growth traits such as stem height, diameter and volume (Gauzere *et al.* 2016). Additionally, SG was clearly reduced by the drought treatment (Fig. 3-8), indicating a negative effect caused by water shortage. This is consistent with other studies showing that under drought conditions, beech saplings show a reduction not only in diameter increment but also in height (Thiel *et al.* 2014). Furthermore, ring widths have been seen to be negatively affected by soil water deficit (Bouriaud *et al.* 2004; Lebourgeois *et al.* 2005). Reduction in stem growth under drought conditions is explained by a shift on carbon allocation priorities, leading to an increment in the root/shoot ratio to facilitate access to soil water (Leuschner *et al.* 2001; Rose *et al.* 2009). Interestingly, the reduction in SG by the drought treatment was more pronounced in 2014. This could be an effect of the delayed growth response of European beech to drought, resulting in a more pronounced reduction in growth the year after the drought occurs (Bolte *et al.* 2007).

Even tough European beech is able to grow on many types of soils, its optimal growth is reached in humid calcareous soils (Jahn 1991). In this study, saplings growing on calcareous soil presented higher SG than saplings on acidic soil, especially in 2014 and also in the overall SG 2013-2014 (Fig. 3-8). It is known that soil characteristics influence the amount of water available for plants (Piedallu *et al.* 2013), which in turn affects nutrient uptake (Geßler *et al.* 2007). Therefore, the response and sensitivity of plants to drought conditions will depend on the type of soil. Indeed, the effect of soil on SG was more noticeable under the drought treatment (Figs. 3-8 B and C). Similar results were found by Thiel *et al.* (2014), who found that a sandy soil with lower water storage capacity and lower nutrient availability resulted in a more severe negative impact on the performance of beech under drought conditions compared to loamy soil.

Despite the negative effect of drought on SG, plants from xeric populations showed higher stem growth than plants from mesic populations, indicating that they were less affected by drought (Fig. 3-9). In fact, other studies investigating growth parameters have also found that populations of beech from mesic sites are more sensitive to drought (Weber *et al.* 2013; Thiel *et al.* 2014), indicating that individuals from xeric populations are better able to cope with the environment they grow in.

Association analysis

One approach for the detection of genetic adaptive variation is to find associations between phenotypic traits and allelic variation. However, population structure and relatedness are confounding factors that can lead to false associations. In this study, saplings collected underneath the same adult tree were more related genetically (see chapter 5); thus, it was necessary to account for relatedness for the phenotypic association analysis. The inclusion of relatedness and population structure and reduced effectively the inflation of association signals, as revealed by the quantile-quantile plots (Appendix 3-1). This suggests that the SNPs deviating significantly from the expected *P* values show true associations with the studied phenotypic traits.

No significant associations were found between the studied SNPs and day of bud burst. It is known that altitude plays an important role in time of bud burst in European beech: populations from high elevations flush earlier than populations from low elevations (Vitasse *et al.* 2009), and genetic differences among them have been found (Kraj & Sztorc 2009). Likewise, SNPs significantly associated with bud burst have been detected in populations in elevation gradients (Müller *et al.* 2015a). However, since the populations selected in this study are located at similar altitudes, selection acting on bud burst could be unlikely or too weak to be detected by the phenotypic association analysis.

In contrast, the phenotypic traits PI_{abs}, PI_{tot} and SG showed significant associations with 5 SNPs, while F_V/F_M showed close to significant associations with 2 SNPs (Table 3-4). The phenotypic variation explained by the SNPs was relatively high ($4.1 \leq R^2 \leq 13.4$) compared to other studies reporting R^2 values between 2.1-6.9 (Hao *et al.* 2012; Porth *et al.* 2013; Müller *et al.* 2015a). Interestingly, the SNP 50_39 in the gene CTR/DRE transcription factor was associated with both chlorophyll fluorescence and growth traits. Heterozygous individuals at this SNP showed a better performance under drought/acidic conditions (Figs. 3-11 B, C and D), indicating overdominance. Other SNP at the same gene, the SNP 50_232,

showed also association with SG 2013-2014 and it seems to have a dominant mode of action, with the allele A conferring higher SG under drought/acidic conditions (Fig. 3-11E). Similarly, the synonymous SNPs IDH_1 and IDH_4 in the gene Isocitrate dehydrogenase were significantly associated with SG, and both seem to have a dominant mode of action (Fig. 3-13) with allele C from IDH_1 and allele G from IDH_4 conferring higher SG under drought/calcareous conditions.

The SNP 110_1_293 in the gene Cytosolic class I small heat-shock protein showed significant association with PI_{abs} under control/acidic conditions (Table 3-4). However, it seems that the two alleles at this SNP have low differences in their phenotypic effect (Fig. 3-12). Indeed, in forest trees many genes with small phenotypic effect control complex traits (Aitken *et al.* 2008), and to detect their effect large sample sizes are required (Hong & Park 2012; Korte & Farlow 2013). Even though the sample size increased when the association analysis was carried out including all saplings, this was not enough to observe a strong effect of the genotype at SNP 110_1_293 on PI_{abs} (Fig. 3-10C). Similarly, the SNP 7_520 showed close to significant association with F_V/F_M in the analysis including all saplings (Table 3-4), but the two alleles at this SNP show low differences in their phenotypic effect (Fig. 3-10B). Thus, a further exploration of the significant association of these SNPs and their effect on PI_{abs} and F_V/F_M will require larger sample sizes.

SNP PP2C_315 also showed close to significant association with F_V/F_M, and individuals with the genotype CC showed better performance than heterozygous (Figs. 3-10A and 3-11A). However, since only one homozygous GG was found, it is not possible to determine the allelic mode of action for this SNP. Further exploration with larger sample size could help get insight into the validity of this association and the likely mode of action for this SNP.

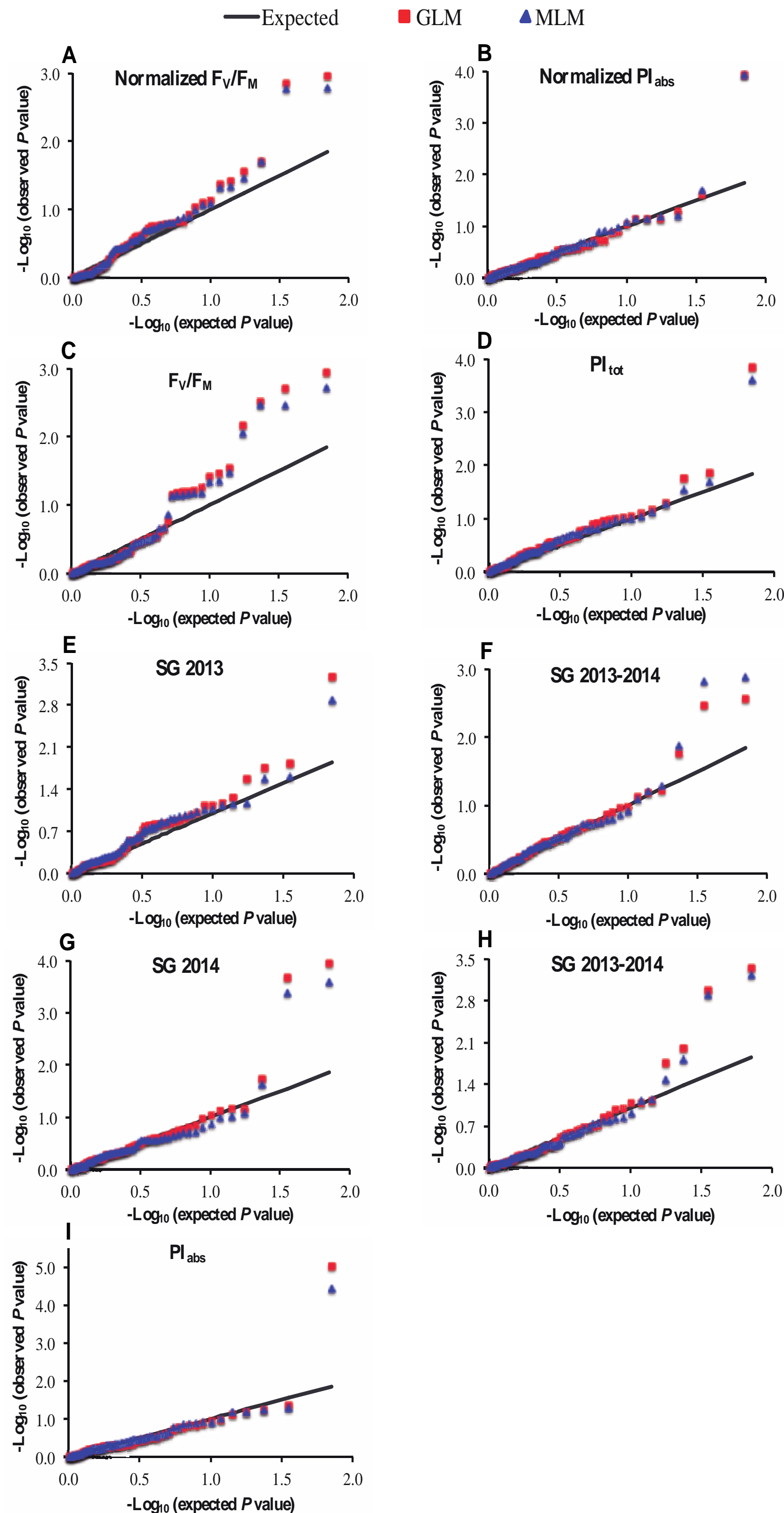
The majority of the SNPs showing association are located in coding regions, representing both non-synonymous and synonymous substitutions (Table 3-4). Traditionally, SNPs in coding regions, and particularly non-synonymous SNPs, are thought to be the main target of natural selection because they cause a change in the aminoacid and thus, can cause a change in the phenotype. Nevertheless, some studies indicate that synonymous substitutions are also important, since they affect mRNA splicing, stability, translation kinetics and ultimately, the production of the protein (Pagani *et al.* 2005; Chamary *et al.* 2006; Komar 2007). Likewise, non-coding regions of the genome are very important because they are responsible for temporal and spatial regulation of gene expression (Barrett *et al.* 2012).

Therefore, synonymous SNPs and non-coding SNPs could also represent genetic variability underlying phenotypic traits.

In conclusion, since populations from xeric sites showed a better performance under the drought treatment as demonstrated by PI_{tot} and SG, this could indicate that those populations are already adapted to dry environments. This is further supported by the SNPs showing significant association with the phenotypic traits, and clearly some genotypes were able to perform better under drought conditions. This provides strong evidence indicating that selection processes leading to adaptation to drought conditions are occurring.

Appendix 3-1

Quantile-quantile plots of estimated $-Log_{10}P$ for the different phenotypic traits showing association under different experimental conditions. **A** and **B** - analysis including all seedlings; **C**, **D**, **E**, and **F** - analysis including seedlings on drought/acidic soil conditions; **G** and **H** - analysis including seedlings on drought/calcareous soil; **I** - analysis including seedlings on control/acidic soil conditions.



4. Association of genetic variation with environment

Introduction

European beech (*F. sylvatica*) is one of the most important and broadly distributed forest tree species in Europe (Bolte *et al.* 2007). In Switzerland, *F. sylvatica* is the second most important tree species, being predominant in the sub-montane and lower montane range (Weber *et al.* 2010). Similar to other beech species, the distribution of European beech depends mainly on temperature, followed by moisture availability (Fang & Lechowicz 2006). The average annual temperature at the southern limits of its distribution is 13.5 ºC, and at the northern limits is 6.6 ºC. Annual precipitation has an average of 906 mm at the southern limits of distribution, and an average of 1272 mm at the northern limits (Fang & Lechowicz 2006).

The distribution of *F. sylvatica* could be affected by climate change (Kramer *et al.* 2010). In Europe, an increment of 1.3º C in the temperature has been already observed since the last half of the 19th century (Kovats *et al.* 2014). Similarly, the frequency of hot days, tropical nights and heat waves has increased since the last half of the 20th century, whereas cold periods and frost days have been reduced (Kovats *et al.* 2014). Additionally, changes in patterns of precipitation have been observed, leading to the occurrence of more extreme events, such as floods and droughts (Lehner *et al.* 2006; Trenberth 2011). In Switzerland, warmer and drier summers have become more common since the 1970s (Beniston & Goyette 2007). By the end of the 21st century, an increment in minimum and maximum temperatures is predicted, as well as more frequent, intense and longer lasting summer warm periods and heat waves, while the number of cold winter days and nights is projected to decrease (Beniston & Goyette 2007; CH2011 2011). Also, precipitation is expected to decrease in summer, affecting mainly the Alpine region, where dry conditions will likely occur (CH2011 2011). Thus, under climate change, the distribution of beech is expected to be affected, with a population reduction in the south and expansion in the north, and a shift in distribution towards higher elevations (Kramer *et al.* 2010; CH2014-Impacts 2014).

The environment is one of the major forces behind natural selection (Rellstab *et al.* 2015). Thus, the most efficient approaches to detect adaptive genetic variation are based on the identification of associations between allele frequencies and environmental variables (Rellstab *et al.* 2015; Stephan 2016). This is the goal of environmental association analyses (EAA), which expect that alleles in a locus under selection and affected by a particular environmental factor might demonstrate a change in allele frequency following

environmental change, for instance, following an environmental gradient (Holderegger *et al.* 2010). An advantage of EAA over other approaches to detect selection, such as outliers tests, is the direct incorporation of environmental variables assumed to be responsible for selection (Schoville *et al.* 2012). Furthermore, EAA are more sensitive to detect subtle changes in allele frequencies caused by weak selection, as in the case of polygenic traits or under high gene flow (Stephan 2016).

Water availability and temperature are among the most important environmental factors affecting plants' survival, and thus, adaptation. By using EAA, it has been possible to detect genetic variability associated with temperature and precipitation in different species, such as *Quercus lobata* (Sork *et al.* 2010), *Arabis alpina* (Poncet *et al.* 2010; Manel *et al.* 2010), *Pinus taeda* (Eckert *et al.* 2010b; a), *P. pinaster* and *P. halepensis* (Grivet *et al.* 2011). Likewise, in *F. sylvatica*, genetic variability at AFLP markers has been associated with temperature (Jump *et al.* 2006) and water availability (Pluess & Weber 2012). More recently, SNPs in candidate genes that might be under climate induced selection have been found (Lalagüe *et al.* 2014; Csilléry *et al.* 2014), and their association with environmental variables such as temperature, precipitation and drought has been determined (Pluess *et al.* 2016). However, the genetic variability underlying adaptation to different environmental conditions in *F. sylvatica* remains insufficiently studied.

A significant increase in the frequency and intensity of summer droughts is predicted under a future climate change scenario. Therefore, the identification of adaptive genetic variability underlying drought tolerance in *F. sylvatica* is of great interest. Thus, the objective of this section is to identify associations between SNPs in climate-related candidate genes supposedly involved in drought tolerance and affected by environmental variables, such as temperature, precipitation and humidity, in populations of *F. sylvatica* occurring in precipitation gradients.

Materials and methods

Plant material

Populations of *F. sylvatica* occurring at two precipitation gradients in Switzerland were selected. Six populations were selected in the Rhine valley with an annual precipitation 849-1334 mm, and six populations were also selected in the Rhone valley with an annual precipitation 603-1012 mm. Leaves from 25 adult trees and 64 saplings per population were

collected, resulting in a total of 300 adult trees and 755 saplings. Leaves were dehydrated with silica gel and stored at room temperature.

DNA isolation

Extraction of DNA from dry leaves was done using the DNeasyTM 96 Plant Kit (Qiagen, Hilden, Germany). Electrophoresis in agarose gel at 1% and 1X TAE as running buffer was carried out to determine the amount and quality of DNA. Before visualization with UV, DNA was stained with Roti®-Safe GelStain (Roth, Karlsruhe, Germany), and compared with a Lambda DNA size ladder (Roche, Mannheim, Germany).

Candidate genes and SNPs

SNPs in candidate genes involved in phenology and stress response have been reported for *F. sylvatica* (Seifert *et al.* 2012; Lalagüe *et al.* 2014; Müller *et al.* 2015b). From those studies, 24 candidate genes were selected, and linkage disequilibrium blocks were identified within each gene using the software htSNPer 1.0 (Ding *et al.* 2005) with the aim of selecting the smallest subset of SNPs characterizing the variability of the gene (Tag SNPs) for posterior genotyping. SNPs also showing signatures of natural selection in previous studies (Csilléry *et al.* 2014; Müller *et al.* 2015a) were also selected. Twenty-one non-synonymous SNPs, 27 synonymous SNPs and 28 non-coding SNPs, for a total of 76 SNPs in 24 genes, were selected for genotyping (Table 4-1). Sequences surrounding the selected SNPs were sent to LGC Genomics Ltd. for primer design and SNP genotyping using the PCR-based KASP™ genotyping assay (Hoddesdon, UK).

Environmental data

Information on climatic variables collected from meteorological stations located near the populations was downloaded from the website of the Federal Office of Meteorology and Climatology MeteoSwiss. Climate normals for the reference period 1961-1990 were used as a proxy for the climate that imposed selection pressure on the early life stages of adult trees, whereas climate normals for the reference period 1981-2010 were used for the saplings. The environmental variables included data on annual and growing season (May-September) temperature and precipitation, as well as heat and summer days (Table 4-2).

Table 4-1 Candidate genes and characteristics of the selected SNPs

Gene	SNP name	Type	Reference
Aldehyde dehydrogenase	ALDH_1	Non-coding	Seifert *et al.* 2012
	ALDH_2	Non-Synonymous	
	ALDH_3	Non-Synonymous	
	ALDH_4	Synonymous	
Isocitrate dehydrogenase	IDH_1	Synonymous	
	IDH_3	Non-coding	
	IDH_4	Synonymous	
Ascorbate peroxidase	APX1_1	Synonymous	
	APX1_2	Non-coding	
	APX4_1	Non-coding	
	APX4_2	Non-Synonymous	
Early responsive to dehydration	ERD	Non-coding	
Dehydrin	Dhn_1	Non-Synonymous	
	Dhn_2	Non-Synonymous	
Glutathione peroxidase	GPX	Non-Synonymous	
Phytochrome B	PhyB	Synonymous	
Cysteine proteinase	CysPro_118	Synonymous	Müller et al. 2015
	CysPro_202	Synonymous	
	CysPro_728	Non-coding	
	CysPro_783	Non-coding	
Chloroplast Chaperonin like	CP10_65	Synonymous	
	CP10_67	Non-Synonymous	
	CP10_377	Non-coding	
	CP10_442	Non-coding	
	CP10_503	Synonymous	
	CP10_749	Synonymous	
	CP10_1317	Non-coding	
	CP10_1428	Non-Synonymous	
Dof zinc finger protein	DAG_81	Non-coding	
	DAG_289	Non-coding	
	DAG_1059	Synonymous	
Histone 3	His3C1_292	Non-coding	
	His3C2_104	Synonymous	
	His3C2_186	Non-coding	
	His3C2_260	Synonymous	
NAC transcription factor	NAC_854	Non-Synonymous	
	NAC_962	Synonymous	
	NAC_1300	Non-coding	
Protein phosphatase 2C	PP2C_315	Non-Synonymous	
	PP2C_391	Synonymous	
	PP2C_791	Non-Synonymous	
	PP2C_941	Non-coding	
	PP2C_1200	Synonymous	
Xyloglucan endotransglucosylase/hydrolase 23	7_258	Non-coding	Lalagüe et al. 2014
	7_520	Non-coding	

Gene	SNP name	Type	Reference
Short chain alcohol dehydrogenase	17_880	Non-coding	
	17_1081	Non-coding	
Potassium transporter 2	39_256	Synonymous	
	39_282	Non-Synonymous	
CRT/DRE binding factor	50_39	Non-Synonymous	
	50_232	Synonymous	
	50_320	Non-coding	
s-adenosyl-l-homocysteine hydrolase	52_1_235	Non-Synonymous	
	52_1_249	Non-Synonymous	
	52_1_368	Synonymous	
Glyceraldehyde 3-phosphate dehydrogenase	68_277	Non-Synonymous	
	68_313	Non-coding	
Light-harvesting complex II protein	88_1_450	Non-coding	
	88_1_727	Synonymous	
	88_1_803	Non-Synonymous	
Catalase	91_2_57	Synonymous	
	91_2_141	Synonymous	
	91_2_231	Synonymous	
	91_2_448	Non-coding	
	91_2_479	Non-coding	
	91_2_504	Non-coding	
1-aminocyclopropane-1-carboxylate oxidase	92_166	Non-coding	
	92_352	Non-Synonymous	
	92_630	Non-coding	
Cytosolic class I small heat-shock protein	110_1_111	Non-Synonymous	
	110_1_293	Synonymous	
	110_1_423	Non-Synonymous	
	110_1_450	Non-Synonymous	
Pectin methylesterase	154_2_137	Synonymous	
	154_2_371	Synonymous	
	154_2_617	Synonymous	

Three derived climatic variables were additionally calculated: potential annual direct incident solar radiation (ASR), the Thornthwaite's moisture index (I_m) (Thornthwaite 1948) and the Ellenberg's climatic quotient (EQ) (Jahn 1991) (Table 4-2). ASR was calculated using data of latitude, slope and aspect according to McCune & Keon (2002). To calculate I_m, first, monthly potential evapotranspiration (PET) according to Thornthwaite (1948) was calculated using the R package SPEI 1.6 (R Core Team 2016). Then, I_m was calculated according to the formula $I_m = \frac{100s-60d}{n}$, where s is the sum of surplus water for the months when precipitation exceeds PET, d is the sum of water deficiency for the months when PET exceeds precipitation, and n is water need (annual PET) (Thornthwaite 1948; Maliva & Missimer 2012). According to Thornthwaite (1948), moist climates have positive values of I_m, and dry climates have negative values. The Ellenberg's climatic quotient (EQ), which is

widely used to describe habitats suitable for the genus *Fagus,* was calculated as $EQ = \frac{\text{Temperature of July (°C)}}{\text{Annual precipitation (mm)}} \times 1000$, (Jahn 1991; Fang & Lechowicz 2006). According to Jahn (1991), regions with values of EQ below 20 represent a pure beech climate, while the beech competitiveness slowly decreases in regions with EQ values between 20-30 and disappears in regions with EQ > 30.

Table 4-2 Abbreviation and description of the geographical and environmental variables

Abbreviation	Description
Lat	Latitude (DD)
Long	Longitude (DD)
MeanAT	Mean Annual Temperature (°C)
MaxAT	Maximum Annual Temperature (°C)
MinAT	Minimum Annual Temperature (°C)
MeanGST	Mean Growing Season[1] Temperature (°C)
MaxGST	Maximum Growing Season[1] Temperature (°C)
MinGST	Minimum Growing Season[1] Temperature (°C)
SD	Summer days[2]
HD	Heat days[3]
AP	Annual Precipitation (mm)
GSP	Growing Season[1] Precipitation (mm)
ADP	Annual Days with Precipitation[4]
GSDP	Growing Season[1] Days with Precipitation[4]
I_m	Thornwaite moisture index
EQ	Ellenmerg's climate quotient (°C/mm)
ASR	Annual Solar Radiation (MJ/cm^2□yr)

[1]From May to September
[2]Number of days with maximum temperature equal to or above 25 °C
[3]Number of days with maximum temperature equal to or above 30 °C
[4]Number of days with precipitation equal or above 1 mm

Information about the environmental variables per population and for the reference periods 1961-1990 and 1981-2010 are presented in Appendix 4-2.

Spearman's rank correlation coefficients among all pairs of environmental variables were calculated. Principal component analysis (PCA) was used to reduce dimensionality of the environmental variables; variables were standardized to a mean of 0 and standard deviation of 1 before PCA analysis. Principal components (PCs) with eigenvalues greater than 1 were kept for the environmental association analysis. All analyses were conducted using the software Statistica 12 (Dell Inc 2015). Climatic PCs were used for further analysis of association with SNP data.

Environmental association analysis

Associations between allelic frequencies and climatic PCs were tested using the R package LEA (Frichot & François 2015). This package tests for associations based on the latent factor mixed models (LFMM), in which associations are tested while estimating the effects of hidden factors, such as population structure and spatial autocorrelation (Frichot *et al.* 2013). After correction for confounding effects, significant association between allele frequencies at a particular locus and environmental variables can be interpreted as evidence for selection (Frichot & François 2015).

A burning period of 5000 and a total number of 10000 cycles were used. Based on the results of the STRUCTURE analysis using the SSR markers (see chapter 2), the number of latent factors K was set to 2 in the saplings and 1 in the adults. Five runs were performed; the z-scores obtained from the different runs were combined using a robust variant of the Stouffer method (Whitlock 2005), and the genomic inflation factor λ (Devlin & Roeder 1999) was computed. *P*-values from the combined z-scores calibrated by λ were obtained as described in the manual of LEA. To ensure that the distribution of *P*-values was suitable for the application of the FDR algorithms, histograms of the *P*-values were obtained, and, if necessary, *P*-values were calibrated by trying different values of λ (François *et al.* 2016). When the histograms showed that the *p*-values were uniformly distributed (Appendix 4-1) (François *et al.* 2016), the Benjamini-Hochberg procedure (Benjamini & Hochberg 1995) with an expected FDR equalled to 10% was used to correct the *P*-values for multiple testing.

Results

Environmental data

Latitude was strongly positively correlated with longitude, minimum temperatures, precipitation variables and the moisture index I_m, and moderately negatively correlated with maximum temperatures, SD, HD and EQ. Longitude had either no correlation or weak positive correlations with most of the variables, most of which were not significant. Maximum temperatures were strongly and positively correlated with SD and HD, while negatively correlated with minimum temperatures and precipitation variables. The Thornthwaite's moisture index I_m was strongly negatively correlated with maximum temperatures and SD and HD, and strongly positively correlated with precipitation. In contrast, the EQ index was positively correlated with maximum temperatures and SD and

HD, and negatively correlated with minimum temperatures and precipitation. ASR had either weak or no correlation with all the environmental variables (Fig. 4-1).

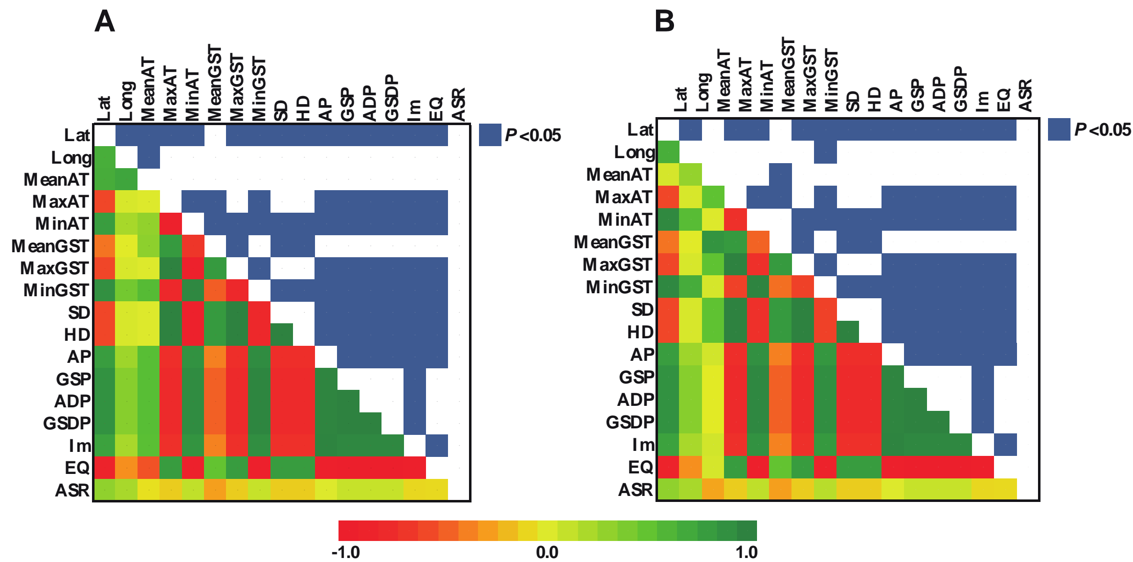


Fig. 4-1 The Spearman's rank correlation coefficients matrix between environmental variables for the reference period **A,** 1961-1990 and **B,** 1981-2010.

The PCA showed that the top three PCs captured the most of the overall variance of geographical and environmental variables for both reference periods: 95.54% for 1961-1990, and 95.99% for 1981-2010 (Table 4-3). These PCs had eigenvalues higher than 1, and they will be referred to as climatic PCs in the rest of the chapter.

Table 4-3 Eigenvalue and variance explained (VE, %) for the first three climatic principal components (PCs) for the reference periods 1961-1990 and 1981-2010

1961-1990			1981-2010		
PC	**Eigenvalue**	**VE, %**	**PC**	**Eigenvalue**	**VE, %**
1	12.310	72.411	1	12.302	72.364
2	2.626	15.446	2	2.793	16.430
3	1.306	7.683	3	1.224	7.203

For both reference periods, the first climatic PC was strongly and positively correlated with latitude, minimum temperatures, precipitation variables and the moisture index I_m, whereas negatively correlated to maximum temperatures, SD, HD and the index EQ (Table 4-4). The second climatic PC was strongly correlated only with mean annual temperature, and the third climatic PC was strongly and positively correlated with solar radiation (Table 4-4).

Table 4-4 Correlation coefficients between the environmental variables and PCs with eigenvalues > 1 for the reference periods 1961-1990 and 1981-2010.

1961-1990				**1981-2010**			
Variable	**PC1**	**PC2**	**PC3**	**Variable**	**PC1**	**PC2**	**PC3**
Lat	**0.801**	0.485	0.314	Lat	**0.814**	-0.462	0.305
Long	0.659	0.580	0.391	Long	0.680	-0.543	0.383
MeanAT	0.066	**0.992**	0.102	MeanAT	-0.258	**-0.961**	0.054
MaxAT	**-0.950**	0.285	-0.001	MaxAT	**-0.967**	-0.225	-0.010
MinAT	**0.970**	-0.177	-0.042	MinAT	**0.980**	0.061	0.127
MeanGST	-0.698	0.645	-0.210	MeanGST	**-0.849**	-0.477	-0.137
MaxGST	**-0.943**	0.250	-0.159	MaxGST	**-0.957**	-0.253	-0.080
MinGST	**0.826**	0.487	-0.118	MinGST	**0.823**	-0.534	0.022
SD	**-0.954**	0.237	-0.125	SD	**-0.882**	-0.458	-0.045
HD	**-0.939**	0.318	-0.060	HD	**-0.819**	-0.557	0.078
AP	**0.937**	0.092	-0.252	AP	**0.932**	-0.184	-0.224
GSP	**0.992**	0.105	-0.062	GSP	**0.984**	-0.156	-0.063
ADP	**0.971**	0.046	-0.178	ADP	**0.953**	-0.149	-0.189
GSDP	**0.986**	0.154	-0.027	GSDP	**0.977**	-0.189	-0.069
Im	**0.908**	0.092	-0.301	Im	**0.902**	-0.215	-0.261
EQ	**-0.961**	0.039	0.173	EQ	**-0.961**	0.029	0.176
ASR	0.101	-0.169	**0.848**	ASR	0.127	0.217	**0.859**

Note. Correlation coefficients > |0.8| are highlighted by the bold font. See Table 4-2 for abbreviations.

Population values for the first climatic PC showed that Ardon, Chamoson, Saxon and Martigny have negative values for this climatic PC (Table 4-5), indicating that these populations are characterized by low values in minimum temperatures, low precipitation related variables and low moisture index I_m, whereas they have higher values in maximum temperatures, SD and HD and EQ values, i.e., drier conditions (Appendix 4-2). Population values for climatic PC2 were negative for Felsberg, Chur, Collombey and Ollon in populations of adults, while positive in populations of saplings (Table 4-5); however, in both cases, this indicates that these populations are characterized by low mean annual temperatures, especially Collombey and Ollon (Appendix 4-2). Population values for climatic PC3 showed that Mastrils, Mels, Saxon, Martigny and Collombey have negative values, indicating that these populations received less solar radiation, when compared to other populations (Table 4-5, Appendix 4-2).

Table 4-5 Population values for the first three climatic principal components (PCs) in a principal component analysis (PCA) of 17 geographical and environmental variables.

Adults	PC1	PC2	PC3	Saplings	PC1	PC2	PC3
Rhine				**Rhine**			
Felsberg	0.155	-0.156	2.078	Felsberg	0.658	0.436	1.868
Chur	0.141	-0.028	1.362	Chur	0.637	0.285	1.119
Malans	2.994	1.122	0.616	Malans	2.817	-1.091	0.725
Mastrils	2.927	1.299	-0.806	Mastrils	2.735	-1.318	-0.762
Sargans	4.040	1.274	0.291	Sargans	3.937	-1.456	0.403
Mels	3.978	1.461	-1.136	Mels	3.859	-1.689	-1.093
Rhone				**Rhone**			
Ardon	-4.707	0.268	0.832	Ardon	-4.788	-0.205	1.003
Chamoson	-4.734	0.317	0.358	Chamoson	-4.823	-0.269	0.505
Saxon	-4.825	0.441	-0.977	Saxon	-4.928	-0.440	-0.891
Martigny	-3.437	0.554	-1.688	Martigny	-3.516	-0.819	-1.615
Collombey	1.698	-3.215	-1.071	Collombey	1.664	3.199	-1.266
Ollon	1.771	-3.338	0.142	Ollon	1.748	3.367	0.002

Environmental association analysis

From the 76 selected SNPs for genotyping, 6 were monomorphic (APX1_2, PhyB, 50_320, 52_1_249, 92_166, 110_1_111); thus, the remaining 70 were used for the association analysis. In total, 24 SNPs (34.3%) showed significant association with at least one of climatic PCs: 6 SNPs in the saplings (8.6%) and 22 SNPs in the adults (31.4%) with 4 of them being common in both saplings and adults - ALDH_1, ALDH_2, 7_258 and 154_2_137 (Table 4-6). Nine of the significantly associated SNPs were non-synonymous (37.5%), 6 synonymous (25%) and 9 non-coding (37.5%). Three SNPs (IDH_3, NAC-854 and 92_630) were associated with more than one of climatic PCs. Overall, SNPs in 17 genes (70.8%) showed significant association, 3 of them in both saplings and adults (ALDH, XTH and PME).

Eight of the SNPs showing association with climatic PCs were also significant in the association analysis with phenotypic traits and/or in the outlier analysis (Table 4-6). Some of these SNPs showed strong differences in allele frequencies, such as IDH_1 and IDH_4. These SNPs are located in the same gene and were also in LD. The frequency of the minor allele (MAF) at these SNPs was lower in populations with negative values of PC1 (mean MAF equalled 0.320 for IDH_1 and 0.318 for IDH_4) when compared to populations with the highest positive values of PC1 (mean MAF equalled 0.505 for IDH_1 and 0.510 for IDH_4) (Figs. 4-2 A and B).

Table 4-6 List of SNPs that significantly correlated with climatic PCs

Gene	SNP	SNP type	Saplings			Adults		
			PC1	PC2	PC3	PC1	PC2	PC3
ALDH	ALDH_1	Non-coding		**			**	
	ALDH_2	Non-Synonymous		**			***	
IDH	**IDH_1**	Synonymous				***		
	IDH_3	Non-coding				**	***	
	IDH_4	Synonymous				***		
APX	APX4_2	Non-Synonymous				**		
ERD	**ERD**	Non-coding						**
CP10	CP10_1317	Non-coding						***
	CP10_1428	Non-Synonymous					***	
CysPro	CysPro_783	Non-coding					**	
DAG	**DAG_1059**	Synonymous					***	
His3	His3C2_186	Non-coding					**	
NAC	**NAC_854**	Non-Synonymous	***		***			
XTH	7_258	Non-coding			**	***		
KT2	39_282	Non-Synonymous				**		
SAHH	52_1_235	Non-Synonymous				***		
	52_1_368	Synonymous				***		
GAPDH	68_277	Non-Synonymous						**
CAT	**91_2_141**	Synonymous				**		
	91_2_448	Non-coding					**	
ACC-oxidase	**92_352**	Non-Synonymous				***	**	
	92_630	Non-coding			**			
sHsps	110_1_423	Non-Synonymous				*		
PME	154_2_137	Synonymous	**					**

Note. SNPs in bold were also significant in the association analysis with phenotypic traits and/or in the outlier analysis. *$P<0.05$, **$P<0.01$, ***$P<0.001$.

Likewise, the mean MAF for SNP 92_352 was lower at negative values of the climatic PC1 (MAF = 0.035) and PC2 (MAF = 0.040), and increased in populations with the highest positive values of both climatic PCs (MAF = 0.162) (Figs. 4-2E and 4-3B). On the other hand, the mean MAF for SNP ERD was higher at negative values of the climatic PC3 (MAF = 0.519), and decreased at the highest positive values (MAF = 0.420) (Fig. 4-4A).

For the rest of the SNPs that also showed association with the phenotypic traits and/or were identified as outliers, changes in MAF were subtler. Mean MAF for NAC_854 was 0.207 at the negative values of the climatic PC1, and 0.212 at the highest positive values (Fig. 4-2C); likewise, mean MAF for this SNP was 0.166 at the negative values of the climatic PC3, and increased slightly to 0.195 at the positive values (Fig. 4-4B). Similarly, mean MAF for 91_2_141 decreased slightly from 0.351 to 0.328 when comparing the most contrasting values of the climatic PC1 (Fig. 4-2D), while mean MAF for DAG_1059 increased slightly from 0.084 to 0.1 between contrasting values of the climatic PC2 (Fig. 4-3A). Mean MAF at

the negative values of the climatic PC3 for 92_630 was 0.486 and slightly decreased to 0.457 in the highest values of this climatic PC (Fig. 4-3C).

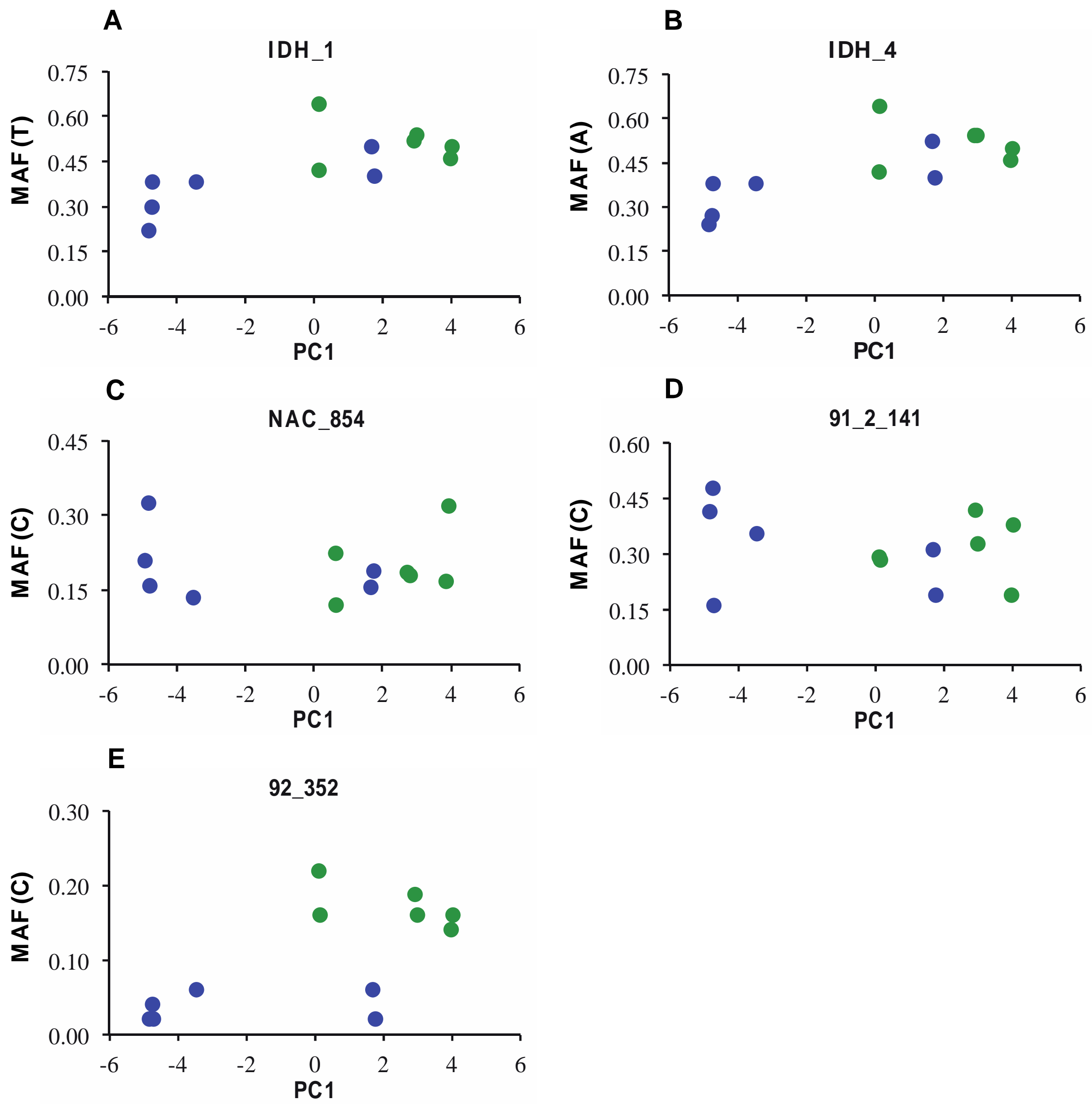


Fig. 4-2 Relationship between climatic variation at the PC1 and minor allele frequencies (MAF) for SNPs that were also identified as outliers. Minor allele is embraced in parenthesis; colors denote regions (green for Rhine and blue for Rhone populations).

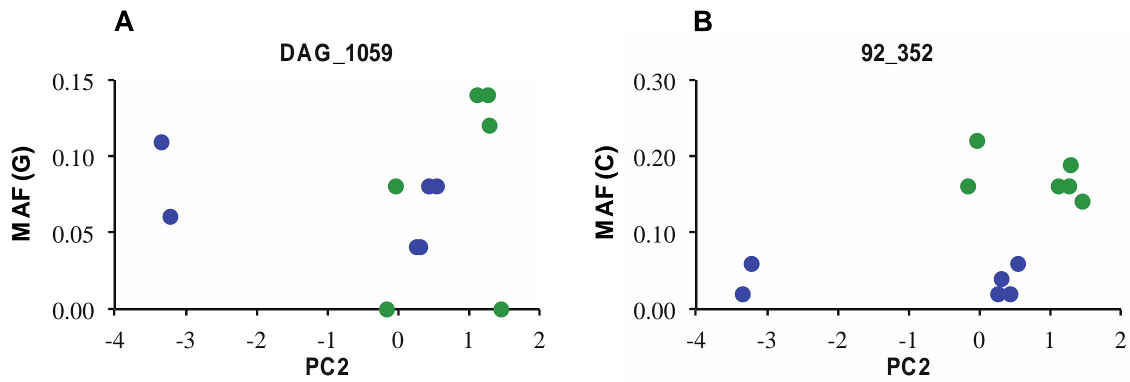


Fig. 4-3 Relationship between climatic variation at the PC2 and minor allele frequencies (MAF) for SNPs that were also identified as outliers. Minor allele is embraced in parenthesis; colors denote regions (green for Rhine and blue for Rhone populations).

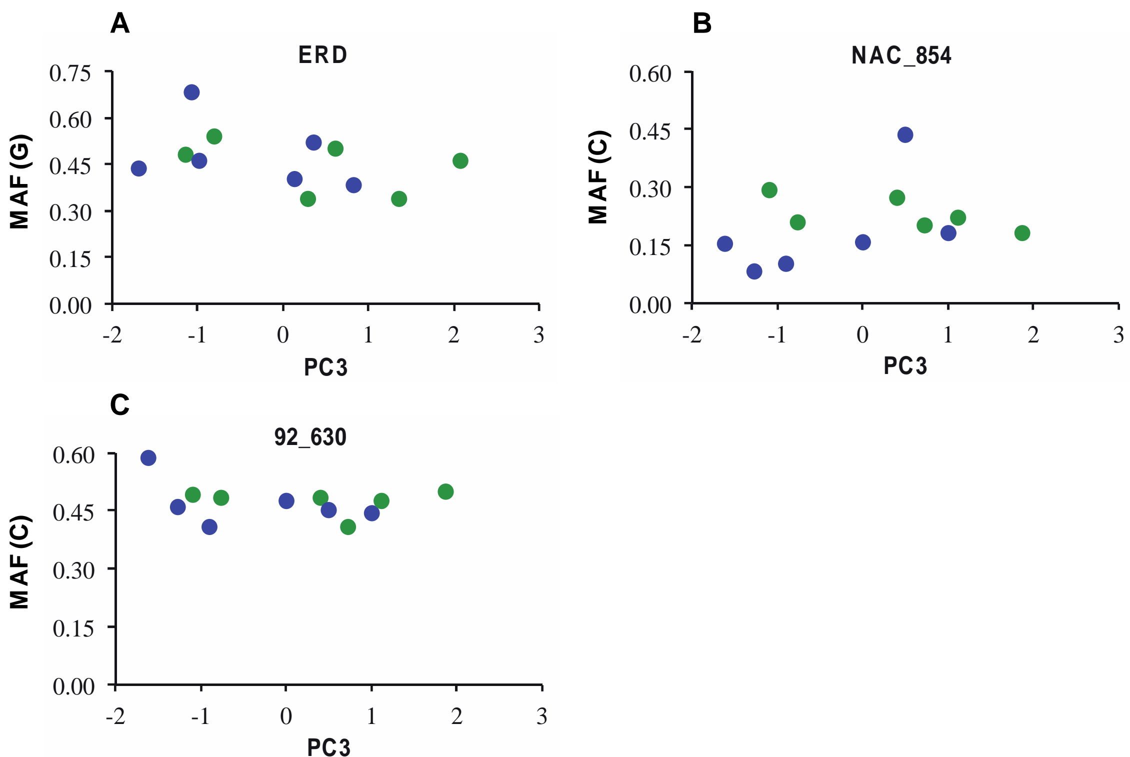


Fig. 4-4 Relationship between climatic variation at the PC3 and minor allele frequencies (MAF) for SNPs that were also identified as outliers. Minor allele is embraced in parenthesis; colors denote regions (green for Rhine and blue for Rhone populations).

Discussion

In this study, environmental variables related to temperature, precipitation and humidity were used to detect genetic variation in SNPs in candidate genes that may be involved in local adaptation in *F. sylvatica*. In total, 24 out of 70 (34.3%) SNPs in 17 out of 24 (70.8%)

candidate genes used in this study showed significant association with climatic PCs (Table 4-6). Not all SNPs in the same gene showed significant association with climatic PCs. The genes included a wide range of cellular functions, including oxidoreductases, hydrolases, oxidases, transferases, transporters, chaperones and transcription factors.

However, approaches to detect genetic adaptive variation are usually prone to false positives, if there is a hidden population structure or spatial autocorrelation (Rellstab *et al.* 2015). Although the LFMM method used in this study incorporates the effect of neutral genetic structure (Frichot *et al.* 2013; Frichot & François 2015), it is recommended to compare results from different approaches, such as EAA and outlier analysis. In this study, 8 (11.43%) out of the 70 studied SNPs were identified by both approaches in 6 (25%) of the studied genes, making them very likely to be involved in local adaptation. This percentages are similar to the ones reported by other studies on beech: Pluess *et al.* (2016) found that 11% of the SNPs in 20% of the genes showed association with environmental predictors.

Not only non-synonymous SNPs were found to be associated with the climatic PCs, but also synonymous SNPs and non-coding SNPs. Since non-synonymous SNPs represent amino acid replacements and thus, a change in protein sequence, they have been traditionally thought to be the main target of natural selection. However, some studies indicate that synonymous substitutions may affect mRNA splicing, stability and translation kinetics (Chamary *et al.* 2006; Komar 2007), and thus, affect the production of the final protein (Pagani *et al.* 2005). Similarly, SNPs in non-coding regions may also be involved in control of gene expression (Barrett *et al.* 2012). Therefore, synonymous and non-coding SNPs can also be subjected to natural selection directly, and not only due to a tight linkage with selective loci.

Some of the SNPs showed strong differences in allele frequencies in contrasting environments. For example, both SNPs in the IDH gene, IDH_1 and IDH_4, showed strong differences in allele frequencies in contrasting environments (Figs. 4-2 A and B). In both SNPs, MAF was lower in populations with negative values for climatic PC1, i.e., drier conditions, high maximum temperatures, high number of SD and HD, low precipitation and low humidity. It means that the alternate alleles (C for IDH_1 and G for IDH_4) had a higher frequency under such environmental conditions. The significant association of these two SNPs with stem growth in the drought experiment (see chapter 3 Fig. 3-13) suggested that alleles C and G provide better performance under drought conditions. This is a strong evidence for the involvement of genetic variability at gene IDH in the local adaptation of the

studied populations of beech. IDH participates in the response to the nitro-oxidative stress, and its expression is induced by salt and drought stress (Liu *et al.* 2010; Leterrier *et al.* 2012).

The 92_352 SNP also showed strong differences in allele frequencies across different environments. MAF at this SNP was also lower at drier conditions and higher mean annual temperatures, demonstrated by negative values in the climatic PC1 and PC2 (Figs. 4-2E and 4-3B). The differences in allele frequencies at this SNP also followed a regional pattern: populations from the Rhine valley have a higher frequency of this allele, than populations from the Rhone valley (Figs. 4-2E and 4-3B), which could be explained by particular environmental conditions in each valley that were not accounted for. Another SNP at the same gene, 92_630, was found significantly associated with the climatic PC3, and showed a slight reduction at populations with high ASR. This SNP has also shown association with temperature, precipitation and drought in populations of beech occurring in Switzerland in a recently published independent study (Pluess *et al.* 2016). Both SNPs belong to the gene ACC-oxidase, a oxidoreductase, whose expression is down-regulated by salt, drought, oxidative stress and ABA (Chen *et al.* 2014).

The minor allele at the ERD SNP decreased in frequency in populations with positive values at the PC3, representing increased amount in solar radiation. ERD is a gene induced by dehydration, and contains cis-elements for the binding of the NAC transcription factors (Shinozaki & Yamaguchi-Shinozaki 2007). The NAC_854 SNP in the NAC gene showed significant association with the PC1 and PC3. However, changes in allele frequencies between populations from contrasting environments were subtle, with MAF increasing slightly in more humid conditions and higher ASR (Figs. 4-2C and 4-4B). This SNP has been associated with bud burst (Müller *et al.* 2015a), a phenological trait that is expected to be affected by rising temperatures under climate change (Schröder *et al.* 2006). Additionally, SNPs in the NAC genes from white spruce (Namroud *et al.* 2008) and boreal black spruce (Prunier *et al.* 2011) have also shown evidence of local adaptation in populations located at different environments. NAC is a transcription factor, whose expression is induced by drought, salinity and ABA (Shinozaki & Yamaguchi-Shinozaki 2007). Other transcription factor in our study is the DAG gene that plays an important role not only in plant development, but also in the biotic and abiotic stress responses (Noguero *et al.* 2013). The DAG_1059 SNP in this gene showed significant association with the climatic PC2, with mean MAF of DAG_1059 increasing slightly in populations with higher annual mean temperatures (Figs. 4-3A).

The CAT gene codes for the enzyme catalase involved in removing the excess of reactive oxygen species produced under stress conditions, and its activity is increased in plants under stress (Sofo *et al.* 2015). The 91_2_141 SNP in this gene showed significant association with the climatic PC1, although differences in MAF between populations from contrasting environments were small. This SNP has also shown evidence of epistatic selection in populations of beech in France occurring in different environments in a recently published independent study (Csilléry *et al.* 2014).

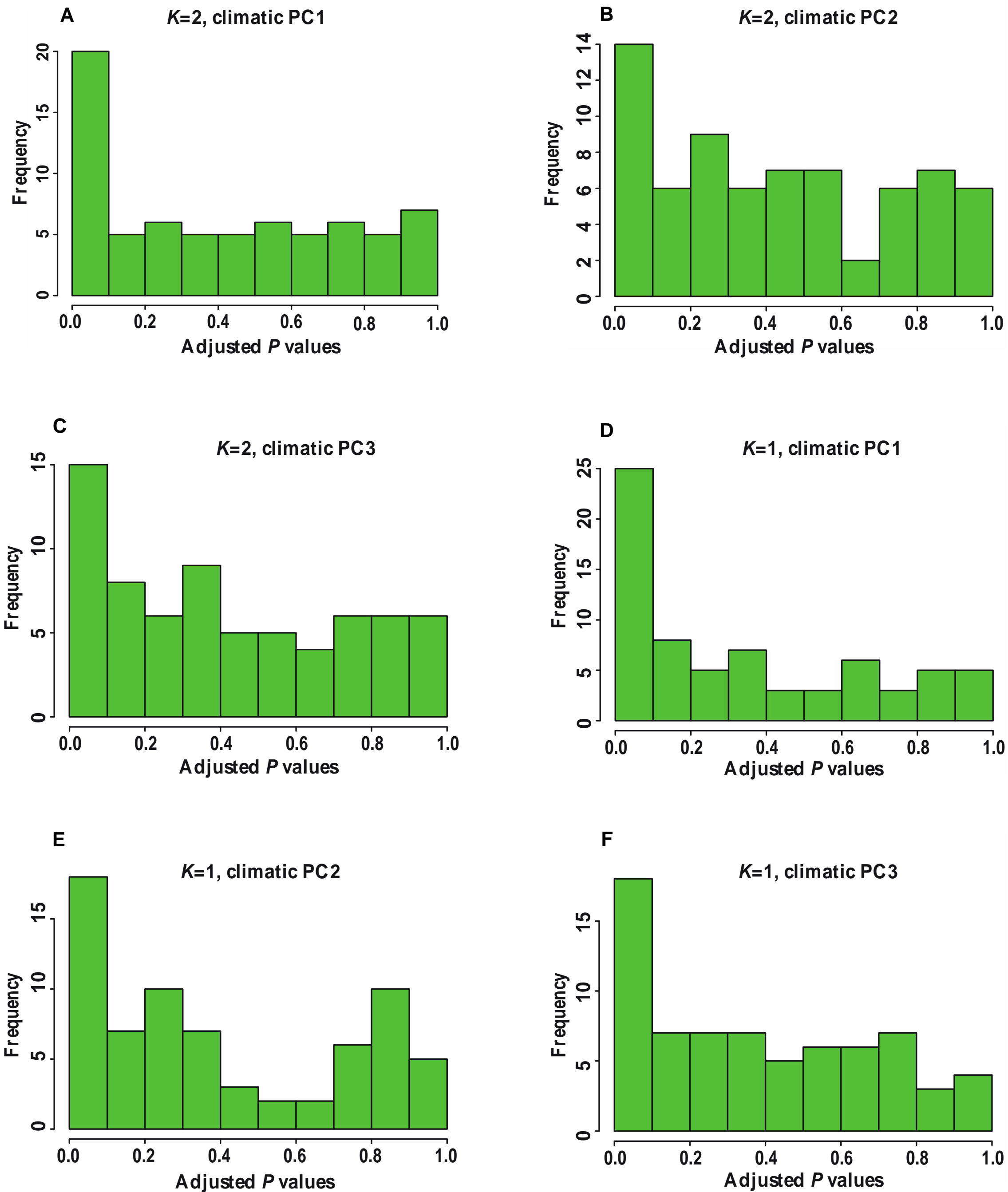
It is possible that other environmental factors that were not accounted for could also explain the differences in allele frequencies observed among populations. In this study, climate data were taken from stations less than 10 km away from the actual populations. However, the Alps have high variability in topography, and climatic factors such as temperature and precipitation can vary over short distances (Baruck *et al.* 2016). Therefore, small-scale heterogeneity and microclimatic conditions particular to each population that were not accounted for, could explain some of the differences in allele frequencies. Furthermore, although precipitation and temperature are the main climatic factors influencing plants' distribution, which is supported by several studies that showed their association with potential adaptive genetic variation in the Alps (Poncet *et al.* 2010; Manel *et al.* 2012; Pluess *et al.* 2016), soil properties might also affect plants' distribution because water availability depends on the interaction between climatic variables and soil characteristics (Piedallu *et al.* 2013). For example, Gärtner *et al.* (2008) found that lower humidity can be compensated for by greater available soil water storage capacity (ASWSC), and thus, allow the grow of beech. Low soil water availability affects survival and competitive interactions between beech and other species (Fotelli *et al.* 2002, 2004) and determines the transition from beech to *Quercus pubescens*, a more drought tolerant tree species (Gärtner *et al.* 2008). In the Alps, soil properties affect not only the present distribution of plants, but also determined the migration pathways during the post-glacial recolonization (Alvarez *et al.* 2009). Furthermore, as the results from the drought experiment carried out in this study suggest, sensitivity to drought depends on soil characteristics (see Chapter 3). Thus, the identification of adaptive genetic variation might be improved by including not only climatic variables but also soil characteristics and microclimatic conditions. However, characteristics of alpine soils vary considerably over short spatial ranges, and soil information is still limited (Baruck *et al.* 2016).

In conclusion, by combining genetic variation in SNPs in candidate genes and climatic environmental data, it was possible to identify loci showing adaptive responses. This opens

new perspectives for understanding the genetic basis of adaptation of *F. sylvatica* to different environmental conditions.

Appendix 4-1

Histograms of adjusted P-values after calibration using the genomic inflation factor λ for saplings (**A, B** and **C**) and adults (**D, E** and **F**). K – number of subpopulations based on the STRUCTURE results.



Appendix 4-2

Appendix 4-2A Data for the environmental variables corresponding to the reference period 1961-1990

Population	Latitude	Longitude	MeanAT, °C	MaxAT, °C	MinAT, °C	MeanGST, °C	MaxGST, °C	MinGST, °C	SD, days
Felsberg	46.854	9.487	9.2	14.2	4.6	15.8	21.3	10.5	40.0
Chur	46.863	9.548	9.2	14.2	4.6	15.8	21.3	10.5	40.0
Malans	46.986	9.570	9.3	13.7	4.9	16.0	21.1	11.1	36.6
Mastrils	46.970	9.543	9.3	13.7	4.9	16.0	21.1	11.1	36.6
Sargans	47.056	9.444	9.3	13.7	4.9	16.0	21.1	11.1	36.6
Mels	47.053	9.411	9.3	13.7	4.9	16.0	21.1	11.1	36.6
Ardon	46.220	7.246	9.2	15.0	4.3	16.5	23.1	10.4	55.3
Chamoson	46.212	7.214	9.2	15.0	4.3	16.5	23.1	10.4	55.3
Saxon	46.146	7.191	9.2	15.0	4.3	16.5	23.1	10.4	55.3
Martigny	46.104	7.108	9.2	15.0	4.3	16.5	23.1	10.4	55.3
Collombey	46.272	6.933	8.9	13.5	4.9	15.5	20.9	10.5	35.1
Ollon	46.303	6.997	8.9	13.5	4.9	15.5	20.9	10.5	35.1

Population	HD, days	AP, mm	GSP, mm	ADP, days	GSDP, days	Im	EQ, °C/mm	ASR, MJ/cm^2•yr
Felsberg	5.4	798.0	430.0	105.2	52.1	34.0	22.6	0.9
Chur	5.4	798.0	430.0	105.2	52.1	34.0	22.6	0.7
Malans	3.5	1095.0	583.0	127.5	61.6	77.8	16.6	0.8
Mastrils	3.5	1095.0	583.0	127.5	61.6	77.9	16.6	0.4
Sargans	3.5	1318.0	647.0	142.2	66.3	114.1	13.8	0.9
Mels	3.5	1318.0	647.0	142.2	66.3	114.1	13.8	0.5
Ardon	10.8	598.0	233.0	82.6	36.2	14.2	31.9	0.9
Chamoson	10.8	598.0	233.0	82.6	36.2	14.2	31.9	0.7
Saxon	10.8	598.0	233.0	82.6	36.2	14.2	31.9	0.4
Martigny	10.8	843.0	319.0	101.2	43.4	47.2	22.7	0.4
Collombey	2.0	1032.0	492.0	122.3	53.6	72.6	17.4	0.5
Ollon	2.0	1032.0	492.0	122.3	53.6	72.6	17.4	0.9

Appendix 4-2B Data for the environmental variables corresponding to the reference period 1981-2010

Population	Latitude	Longitude	MeanAT, °C	MaxAT, °C	MinAT, °C	MeanGST, °C	MaxGST, °C	MinGST, °C	SD, days
Felsberg	46.854	9.487	10.0	15.0	5.6	16.8	22.3	11.7	50.8
Chur	46.863	9.548	10.0	15.0	5.6	16.8	22.3	11.7	50.8
Malans	46.986	9.570	10.1	14.6	5.7	16.9	22.1	12.1	50.3
Mastrils	46.970	9.543	10.1	14.6	5.7	16.9	22.1	12.1	50.3
Sargans	47.056	9.444	10.1	14.6	5.7	16.9	22.1	12.1	50.3
Mels	47.053	9.411	10.1	14.6	5.7	16.9	22.1	12.1	50.3
Ardon	46.220	7.246	10.1	16.0	5.1	17.5	24.1	11.5	68.8
Chamoson	46.212	7.214	10.1	16.0	5.1	17.5	24.1	11.5	68.8
Saxon	46.146	7.191	10.1	16.0	5.1	17.5	24.1	11.5	68.8
Martigny	46.104	7.108	10.1	16.0	5.1	17.5	24.1	11.5	68.8
Collombey	46.272	6.933	9.8	14.5	5.6	16.6	21.9	11.6	41.4
Ollon	46.303	6.997	9.8	14.5	5.6	16.6	21.9	11.6	41.4

Population	HD, days	AP, mm	GSP, mm	ADP, days	GSDP, days	Im	EQ, °C/mm	ASR, ASR, $MJ/cm^2 \cdot yr$
Felsberg	10.3	849.0	466.0	104.6	51.7	35.3	22.5	0.9
Chur	10.3	849.0	466.0	104.6	51.7	35.3	22.5	0.7
Malans	8.7	1114.0	593.0	125.7	60.5	73.4	17.2	0.8
Mastrils	8.7	1114.0	593.0	125.7	60.5	73.4	17.2	0.4
Sargans	8.7	1334.0	672.0	142.5	66.0	107.9	14.4	0.9
Mels	8.7	1334.0	672.0	142.5	66.0	107.9	14.4	0.5
Ardon	16.0	603.0	262.0	82.1	36.9	9.7	33.3	0.9
Chamoson	16.0	603.0	262.0	82.1	36.9	9.7	33.3	0.7
Saxon	16.0	603.0	262.0	82.1	36.9	9.7	33.3	0.4
Martigny	16.0	855.0	343.0	100.1	44.2	42.6	23.5	0.4
Collombey	3.1	1012.0	501.0	117.8	53.1	60.6	18.9	0.5
Ollon	3.1	1012.0	501.0	117.8	53.1	60.6	18.9	0.9

Max – maximum*, Min* – Minimum, *AT*- Annual Temperature, *GST* – Growing Season Temperature, *SD* – Summer Days (equal to or above 25 °C), *HD* – Heat Days (equal to or above 30 °C), *AP* – Annual Precipitation, *GSP* – Growing Season Precipitation, *ADP* – Annual Days with Precipitation, *GSDP* - Growing Season Days with Precipitation, I_m Thornthwaite's moisture index*, EQ* – Ellenberg's Quotient, *ASR* - annual direct incident solar radiatio

5. Spatial genetic structure, relatedness and parental assignment

Introduction

Spatial genetic structure describes spatial distribution of genotypes. At a local or fine spatial scale neighboring individuals could be more related and, respectively, more similar genetically (Vekemans & Hardy 2004). In plants, this is mainly caused by limited dispersal of pollen and seeds (Vekemans & Hardy 2004).

Fagus sylvatica is a dominant forest tree species in Europe. It is a monoecious tree, mainly outcrossing and wind pollinated (von Wuehlisch 2008). Within stands, pollen dispersal distance has been estimated between 37 m (Wang 2004) and 57 m (Oddou-Muratorio *et al.* 2011); however, it has been found that pollen can travel for much longer distances, for hundreds or even thousands of kilometers (Belmonte *et al.* 2008), which is supported by the high pollen migration rates observed in this species (Oddou-Muratorio *et al.* 2011; Piotti *et al.* 2012).

The seeds of *F. sylvatica* are mainly dispersed by gravity, being released under the canopy of the mother tree. A mean seed dispersal distance of 11 m has been estimated (Oddou-Muratorio *et al.* 2011), although there are reports of seed dispersal distances within the range of 4-50 m (Millerón *et al.* 2013; Bontemps *et al.* 2013). While primary dispersal occurs by gravity, the beech nuts are an important food source for several small animals including rodents, nuthatches, great tits and jays, which may also play an important role in seed dispersal by hiding the seeds and failing to retrieve all of them, contributing to dispersal far away from the mother tree (Jensen 1985; Nilsson 1985; Perea *et al.* 2011; Packham *et al.* 2012).

Investigation of local spatial genetic structure in beech stands from Germany has revealed that they form a strong family structure up to 30 m (Vornam *et al.* 2004). Similarly, studies on beech populations in Poland (Chybicki *et al.* 2009) and other European countries (Piotti *et al.* 2013) has revealed significant spatial genetic structure up to 40 and 20 m, respectively.

Relatedness can be defined as the fraction of alleles among individuals that are identical by descent (Blouin 2003). In genomic association analysis, relatedness is considered a confounding factor that could lead to false associations (Sillanpää 2011). This is because related individuals are more likely to be more genetically similar, causing a higher correlation

of the phenotypic trait among individuals from the same family (Foulkes 2009). To overcome this problem, relatedness between pairs of individuals should be estimated and accounted for in phenotypic association analysis (Foulkes 2009; Sillanpää 2011).

In this study, 2-4 saplings under the crown of the same adult tree were collected from 16 adult trees per population; thus, due to the strong spatial genetic structure in *F. sylvatica,* it is very likely that the collected saplings are related. The objective of this chapter is to establish the degree of relatedness between pairs of saplings collected under the same adult tree. Furthermore, since adult trees were also sampled and genotyped, a maternal assignment analysis was performed.

Materials and methods

Plant material

Saplings and adults from 12 populations of *F. sylvatica* located in the Rhone and Rhine valleys in Switzerland were used in this study. In a first stage, 16-31 adult trees about 50 m apart to each other were selected, and 2-4 saplings underneath them with a height of ~20 cm were excavated and transferred in spring 2011 to the WSL institute in Switzerland for a drought experiment. During this stage the leaves from these adult trees were not collected for genotyping. In a second stage, in 2014, using maps and GIS tracking, we searched for the adult trees underneath which the saplings were collected in a first stage. However, since these adult trees were not labeled and some populations had a high density of beech trees, their accurate identification was difficult. Thus, leaves from 25 adult trees about 50 m apart to each other were sampled in each population; some of these trees could correspond to the adult trees where saplings were collected. In total, leaves from the 755 saplings selected in the first stage and from the 300 adult trees selected in the second stage were sampled, dehydrated with silica gel and stored at room temperature until DNA isolation.

DNA isolation and microsatellite loci genotyping

DNA was isolated from dry leaves using the DNeasyTM 96 Plant Kit (Qiagen, Hilden, Germany). The amount and quality of the DNA were examined using electrophoresis in agarose gel at 1% and 1x TAE as running buffer. DNA was stained with Roti®-Safe GelStain (Roth, Karlsruhe, Germany), visualized by UV illumination, and compared with a Lambda DNA size ladder (Roche, Mannheim, Germany).

Individuals were genotyped at 13 microsatellite loci. Ten of them are genomic microsatellites (gSSRs) - six markers were originally developed for *F. sylvatica: FS3-04* (Pastorelli *et al.* 2003), *msf11* (Vornam *et al.* 2004), *csolfagus_06*, *csolfagus_19* (Lefèvre *et al.* 2012), *Fagsyl_002929* and *Fagsyl_003994* (Pluess & Määttänen 2013), and four markers *sfc0018*, *sfc0161*, *sfc1063* and *sfc1143* were originally developed for *F. crenata* (Asuka *et al.* 2004). Other three microsatellite loci *GOT066*, *FIR065* and *FIR004* are EST-linked (EST-SSRs) and were originally developed for *Quercus robur* (Durand *et al.* 2010) and successfully used for *F. sylvatica.*

The polymerase chain reaction (PCR) amplifications were performed using fluorescent dye labeled primers as follows: 6-carboxyfluorescein dye (FAM) for *mfs11*, *sfc0161*, *sfc1063*, *csolfagus_06*, *csolfagus_19*, *Fagsyl_003994* and *FIR004*; and 6-hexachlorofluorescein dye (HEX) for *sfc0018*, *sfc1143*, *Fagsyl_002929*, *GOT066*, *FIR065* and *FS3-04*. This allowed us to assemble four different PCR amplification multiplexes. The first multiplex contained the *FS3-04* and *msf11* markers. The second multiplex contained all four *sfc* markers. The third multiplex contained the *csolfagus* and *Fagsyl* markers. The fourth multiplex contained the EST markers. The PCR amplifications were performed in a total volume of 15 µL containing 2 µL of genomic DNA (about 10 ng), 1x reaction buffer (0.8 M Tris-HCl pH 9.0, 0.2 M $(NH_4)_2SO_4$, 0.2% *w/v* Tween-20; Solis BioDyne, Tartu, Estonia), 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.3 µM of each forward and reverse primer and 1 unit of *Taq* DNA polymerase (HOT FIREPol® DNA Polymerase, Solis BioDyne, Tartu, Estonia). The amplification conditions were as follows: an initial denaturation step at 95 °C for 15 min, followed by 30 cycles consisting of a denaturing step at 94 °C for 1 min, an annealing step at 55 °C (first, second and third multiplexes) or at 47 °C (EST multiplex) for 30 s and an extension step at 72 °C for 1 min. After 30 cycles, a final extension step at 72 °C for 20 min was executed. The PCR fragments were separated on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster City, USA). The GS 500 ROX™ (Applied Biosystems, Foster City, USA) was used as an internal size standard. Genotyping was done using the GeneMapper 4.1® software (Applied Biosystems, Foster City, USA).

Relatedness and spatial genetic structure analysis

Based on microsatellite genotypes Queller & Goodnight (1989) relatedness coefficient r_{QG} was calculated between pairs of saplings within the same population using the GenAlex 6.5 software (Peakall & Smouse 2012). Unrelated individuals are expected to have a $r_{QG} = 0$, half-sibs a $r_{QG} = 0.25$, and full-sibs a $r_{QG} = 0.5$. To test whether r_{QG} coefficients for saplings

collected at the same site (underneath the same adult tree) were significantly different from coefficients for saplings collected in different sites (underneath different adult trees), a Mann-Whitney U test was conducted using the Statistica 12 software (Dell Inc 2015).

In addition, a spatial autocorrelation analysis was performed using the method described in Smouse & Peakall (1999) and implemented in the GenAlex 6.5 software (Peakall & Smouse 2012). To define the upper and lower limits of the 95% confidence interval, 999 permutations were performed. Since from a statistical point of view it is recommendable to keep the number of pairs of individuals compared approximately constant within each distance class (Diniz-Filho *et al.* 2013), analyses were carried out using the option “even sample classes” that distributes as equal number of pairs of individuals per distance class as possible.

Parentage assignment

Based on microsatellite genotypes, a categorical parentage assignment was carried out with the CERVUS software (Marshall *et al.* 1998). Saplings were assigned to a single tentative parent using the maternity analysis option. Running conditions accepted incomplete sampling of candidate mothers and a genotyping error of 1%. A level of confidence of 80% was used, and the critical value of Δ above which parentage is assigned was determined using 100000 offspring simulated.

Results

Relatedness and spatial genetic structure

The mean pairwise values of the relatedness coefficients r_{QG} per population were significantly different between saplings collected at the same site and saplings collected at different sites (Table 5-1). Mean r_{QG} values were very close to zero in saplings collected at different sites (Table 5-1), while higher relatedness values were observed for saplings collected at the same site, between 0.129 (Mastrils and Sargans) and 0.221 (Collombey) (Table 5-1). For all populations, the r_{QG} coefficients close to zero were the most frequent class in saplings collected at both different and the same site, but number of coefficients close to zero was much higher in sapling pairs collected from different sites (Fig. 5-1). Spatial autocorrelation analyses detected significant spatial genetic structure in the first distance class for all populations, extending from 38 m (Felsberg) to 130 m (Saxon) (Fig. 5-2).

Table 5-1 Mean values per population of the relatedness coefficient r_{QG} for pairs of saplings collected at the same site (underneath the same adult tree) and at different sites (underneath different adult trees)

Population	The same site	Different sites
Felsberg	0.197	0.055
Chur	0.191	0.058
Malans	0.156	0.057
Mastrils	0.129	0.062
Sargans	0.129	0.062
Mels	0.147	0.059
Ardon	0.179	0.062
Chamoson	0.205	0.061
Saxon	0.194	0.059
Martigny	0.145	0.059
Collombey	0.221	0.06
Ollon	0.161	0.057
Mean	0.171	0.059

Note. Mann-Whitney U test was significant for all populations at $P < 0.001$

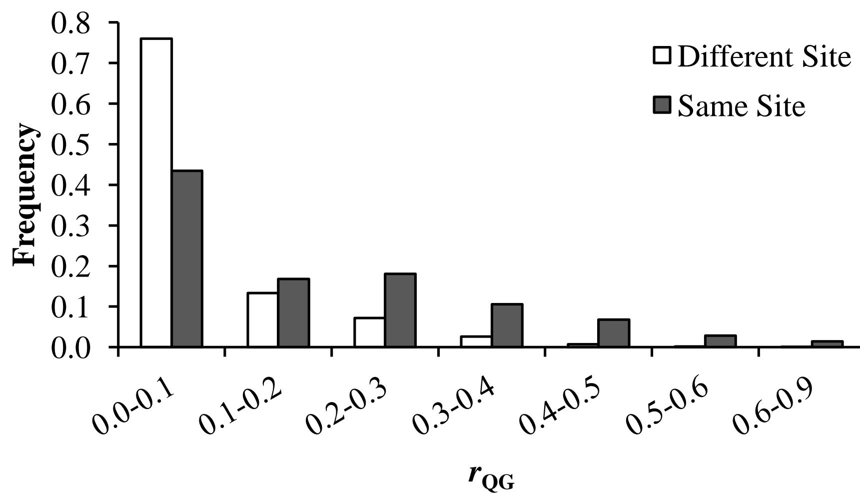


Fig. 5-1 Distribution of pairwise r_{QG} coefficients in saplings collected at different sites (underneath different adult trees) and at the same site (underneath the same adult tree) for all populations.

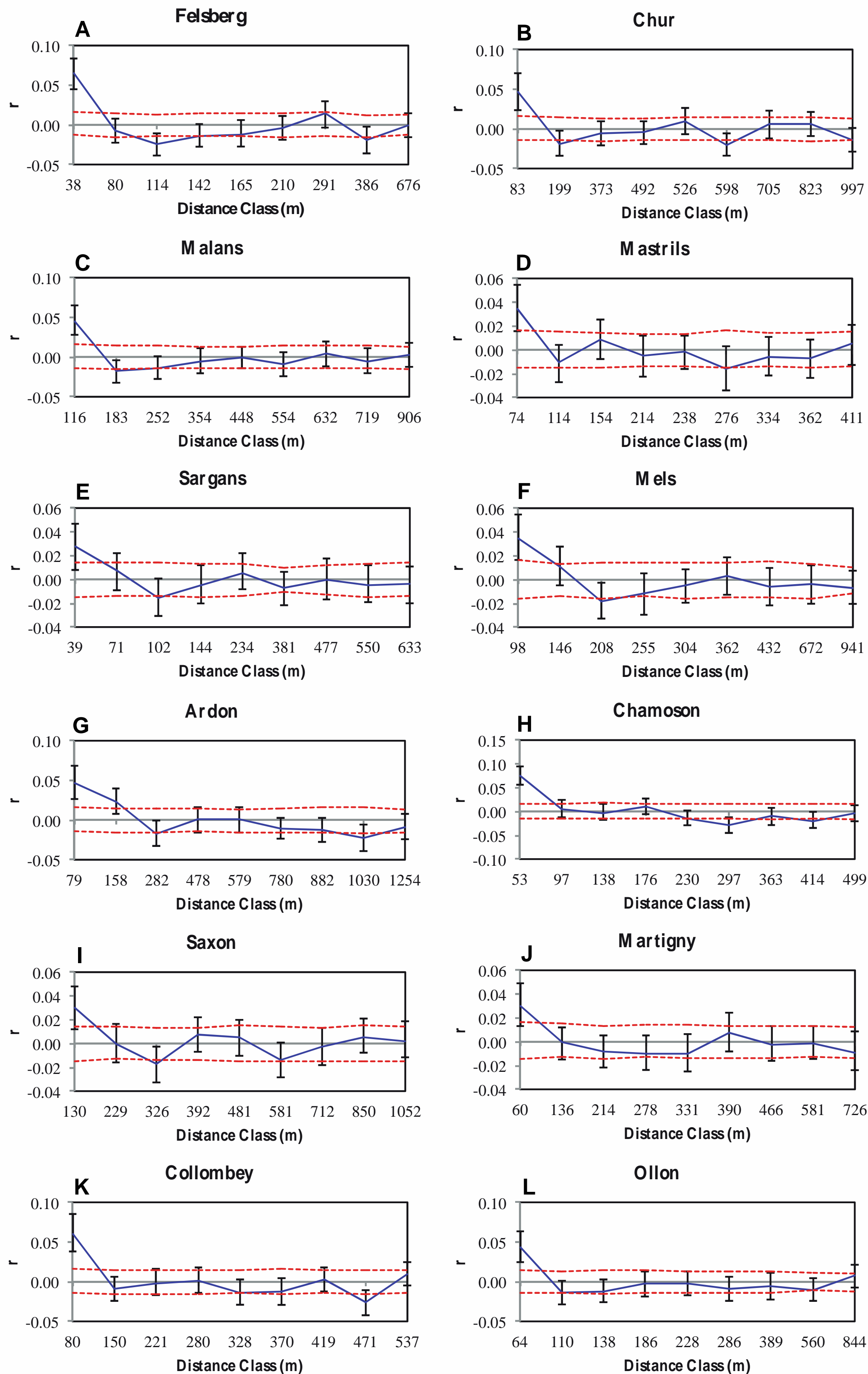


Fig. 5-2 Correlograms showing the correlation coefficient *r* as a function of geographic distance, using even sample size in each class. Red dotted lines represent the 95% confidence interval under a null hypothesis of no spatial structure. Blue solid lines represent the correlation coefficient *r*, and the 95% error bars around mean *r* values generated by bootstrapping are shown.

Parentage assignment

A tentative parent was identified for 44.4% of the saplings. Mastrils was the population with the lowest proportion of successfully assigned saplings (33.9 %), while Chamoson had the highest (62.5 %, Table 5-2). However, if no mismatches in the genotype are allowed in the assignment, the proportion of saplings with a tentative parent assigned was almost twice as less (Table 5-2).

Table 5-2 Percentage of saplings for which a tentative parent was assigned considering no mismatches or 1-2 mismatches in the genotype.

Population	No mismatches	1-2 mismatches	Total
Felsberg	27.4	22.6	50.0
Chur	34.9	7.9	42.8
Malans	31.3	15.6	46.9
Mastrils	9.7	24.2	33.9
Sargans	17.5	20.6	38.1
Mels	20.0	25.0	45.0
Ardon	17.5	20.6	38.1
Chamoson	28.1	34.4	62.5
Saxon	28.1	12.5	40.6
Martigny	25.0	14.1	39.1
Collombey	42.9	15.8	58.7
Ollon	14.3	22.2	36.5
Mean	24.7	19.6	44.4

As it could be expected, saplings collected at the same site were usually assigned to the same tentative parent, which is most likely their seed parent, i.e., mother tree. This was observed in most populations, especially in Collombey (Appendix 5-1K). However, in some cases, saplings collected at the same site were assigned to different tentative parents, for example, in the case of saplings W1-33 to W1-36 (Appendix 5-1G) and saplings W2-10 to W2-12 (Appendix 5-1H). Similarly, in other cases, it was possible to assign a tentative parent tree to only one sapling among those collected at the same site, such as in the case of sapling G1-06 (Appendix 5-1A) and sapling G2-55 (Appendix 5-1B). Interestingly, in almost all populations there were adult trees that were tentative parents for saplings collected at different sites, for example adult trees G1-A1 and G1-A9 (Appendix 5-1A) and G2-A16 and G2-A24 (Appendix 5-1B).

Discussion

Significant spatial genetic structure was found in the saplings within 38-130 m distance (Fig 5-2). This is in line with other studies that have also reported significant spatial autocorrelation up to distances of 20-40 m (Vornam *et al.* 2004; Chybicki *et al.* 2009; Piotti *et al.* 2013) or 110 m (Jump & Peñuelas 2007). The main characteristic of the spatial genetic structure found was a decrease in relatedness with increasing distance among individuals. Higher relatedness measured by the r_{QG} coefficient was detected between pairs of saplings collected at the same site, i.e., under the same adult tree (Table 5-1). This is expected taking into account that in beech, seeds are mostly gravitationally dispersed and distributed around the mother tree (Millerón *et al.* 2013), making it more likely that the saplings collected at the same site had been produced by the same mother tree and thus, more likely to be related. In fact, the relatedness coefficient r_{QG} values for the saplings collected underneath the same adult tree were close to 0.2 (Table 5-1), only slightly lower than 0.25, which is the relatedness coefficient expected for half-sibs (Queller & Goodnight 1989). However, it may be incorrect to assume that saplings growing underneath of an adult tree are growing underneath their mother (Ashley 2010). Indeed, about half of the pairs of saplings collected underneath the same mother tree had relatedness coefficients close to zero (Fig. 5-1), indicating no relatedness (Queller & Goodnight 1989; Hedrick 2005). Furthermore, the parentage analysis showed that, in some cases, the assignment of a tentative parent was possible only for one or some of the saplings collected at the same site. Moreover, some of the adult trees were assigned as tentative parents of saplings collected at different sites (Appendix 5-1). Both results may imply the dispersion of some saplings away from their seed parent, i.e., mother tree. This is consistent with several studies demonstrating that beech saplings are found at longer distances from their mother trees than seeds (Millerón *et al.* 2013; Bontemps *et al.* 2013). For example, Millerón *et al.* (2013) found that while mean seed dispersal distance ranged from 13.1 to 20.1 m, dispersal distances for saplings ranged from 156.2 to 401.2 m. Seed dispersal in beech occurs primarily by gravity, and at sites with significant inclination, it can contribute to dispersion downslope away from the mother tree (Bontemps *et al.* 2013). Besides, secondary dispersal by animals can also occur, particularly by rodents (Jensen 1985) and birds (Nilsson 1985; Perea *et al.* 2011) that can disperse seeds for tens and hundreds of meters. Thus, differences in dispersal distances between seeds and saplings are usually attributed to the combined affect of primary and secondary dispersal (Millerón *et al.* 2013; Bontemps *et al.* 2013). Additionally, it is worth to mention that in this

study the adult trees that were assigned as tentative parents of saplings collected at different sites cannot be ruled out as pollen parents, rather than seed parents.

Dispersed seeds, seedlings and saplings represent a challenge for parentage assignment, because neither parent is really known (Ashley 2010). This challenge is heightened by the longer dispersion of beech saplings away from the mother tree (Millerón *et al.* 2013; Bontemps *et al.* 2013), which makes less likely to find a sapling underneath its mother tree and thus, makes seed parent assignment more difficult. Even if saplings are not dispersed far away, and mother tree identification is possible, as occurred for most of the saplings from Collombey (Appendix 5-1K), in this study the tracking of the adult trees under which saplings were collected in the first stage was challenging, because those adult trees were not labeled, and some populations had a high density of beech. Despite these circumstances, the percentage of saplings successfully assigned, allowing for mismatches, was 44.4%. In other studies with exhaustive sampling of candidate parents, the percentage of saplings successfully assigned ranged between 31 % and 94 % (Millerón *et al.* 2013; Bontemps *et al.* 2013). The differences were attributed to different rates in pollen and seed immigration.

Microsatellite loci are usually the markers of choice for parentage and relatedness analysis. Due to their high polymorphism, the probability that two individuals share an allele by chance is very low (Weir *et al.* 2006). Normally, by using several microsatellite markers, the probability of exclusion is sufficiently high (97-99%) for their use in parentage analysis (Ashley 2010). However, microsatellites have several limitations. Null alleles, allele dropout and mutations may occur, causing mismatches in the genotypes of related individuals and thus, leading to false exclusions (Jones *et al.* 2010; Ashley 2010) They can present null alleles, that are alleles that do not amplify during PCR due to mutations in the primer-annealing site (Varshney *et al.* 2005). Among the microsatellite loci used in this study, seven were transferred from other species of the same family *Fagus crenata* and *Quercus robur*. Microsatellite transference could increase the probability of null alleles due to differences in the sequence of the binding site of the primer, leading to failure in amplification. Although these loci did not show evidence of null alleles (see Chapter 2), miss-genotyping due to null alleles cannot be ruled out.

To conclude, although the sampling design used here was originally aimed neither at studying the spatial distribution of the genetic variation within populations (see, for example, Vornam *et al.* 2004; Piotti *et al.* 2013) nor for parentage analysis (see, for example, Millerón *et al.* 2013; Bontemps *et al.* 2013), some general trends in line with findings in other studies were observed. Firstly, significant local spatial genetic structure was found in the saplings,

meaning that the closer the individuals are located, the more they are related genetically. Thus, it was important to take this information into account for the phenotypic association analysis (see Chapter 3) to avoid false positives (Foulkes 2009; Sillanpää 2011). Secondly, parentage analysis showed that saplings collected underneath the same tree do not necessarily represent progeny of this tree due to seed dispersion by both gravity and animals. Indeed, long distance seed dispersal is important because it allows plant species to colonize suitable habitats, and thus, has a significant impact on survival in a changing environment (Cain *et al.* 2000; Johnson *et al.* 2017). Additionally, gene flow by seed dispersal has the potential of increasing genetic variation, and thus, enhance local adaptation (Kremer *et al.* 2012).

Appendix 5-1

Appendix 5-1A. Parent assignment for saplings from Felsberg. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching	Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
G1-01	G1-A2	na	-	G1-33[f]	G1-A5	G1-A5	0
G1-02		na	-	G1-34		na	-
G1-03	G1-A15	na	-	G1-35		na	-
G1-05		na	-	G1-36[f]		G1-A5	0
G1-06[a]		G1-A1	0	G1-37	nd	G1-A14	2
G1-07	nd	na	-	G1-38		na	-
G1-09	G1-A3	G1-A3	0	G1-39		na	-
G1-10		na	-	G1-40		na	-
G1-45[b]		G1-A17	1	G1-41	G1-A20	na	-
G1-11	G1-A17	G1-A16	2	G1-42		na	-
G1-12[h]		G1-A9	0	G1-43[g]		G1-A6	1
G1-13		na	-	G1-44		na	-
G1-14		na	-	G1-47[h]	G1-A21	G1-A9	0
G1-15	nd	na	-	G1-48[d]		G1-A21	0
G1-16		na	-	G1-49[d]		G1-A21	0
G1-17[c]	nd	G1-A2	0	G1-50[d]		G1-A21	0
G1-18		na	-	G1-51	G1-A8	G1-A13	1
G1-19		na	-	G1-52[b]		G1-A17	1
G1-20[c]		G1-A2	1	G1-53		na	-
G1-21[e]	G1-A6	G1-A4	2	G1-54		na	-
G1-22[g]		G1-A6	0	G1-55	nd	na	-
G1-23[g]		G1-A6	0	G1-56		na	-
G1-24[g]		G1-A6	0	G1-57		na	-
G1-25[a]	G1-A18	G1-A1	0	G1-58		na	-
G1-26		G1-A20	1	G1-59	nd	na	-
G1-27		G1-A7	1	G1-60		G1-A14	2
G1-28		na	-	G1-61		G1-A14	2
G1-29[e]	G1-A7	G1-A4	0	G1-62[b]		G1-A17	1
G1-31		na	-	G1-63	nd	G1-A25	2
G1-32		G1-A22	0	G1-64		G1-A13	0
				G1-65		na	-
				G1-66		na	-

Appendix 5-1B. Parent assignment for saplings from Chur. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching	Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
G2-01	nd	na	-	G2-33	nd	G2-A10	2
G2-02		na	-	G2-34		na	-
G2-03		na	-	G2-35[b]		G2-A16	0
G2-04		na	-	G2-36		na	-
G2-05[a]	G2-A1	G2-A1	0	G2-37	G2-A19	na	-
G2-06[a]		G2-A1	0	G2-38[e]		G2-A19	0
G2-07[a]		G2-A1	0	G2-39[e]		G2-A19	0
G2-08[a]		G2-A1	0	G2-40[e]		G2-A19	1
G2-09	nd	na	-	G2-41	nd	na	-
G2-10		na	-	G2-42		na	-
G2-11		na	-	G2-43		na	-
G2-12		na	-	G2-44		na	-
G2-13	nd	na	-	G2-45	nd	na	-
G2-14		na	-	G2-46[g]		G2-A5	1
G2-15		na	-	G2-47		G2-A4	0
G2-16		na	-	G2-48		na	-
G2-17	G2-A8	na	-	G2-49	G2-25	na	-
G2-18[g]		G2-A5	0	G2-51		na	-
G2-19		na	-	G2-52		na	-
G2-20[f]		G2-A24	0	G2-53	nd	na	-
G2-21	G2-A5	na	-	G2-54		na	-
G2-22[b]		G2-A16	0	G2-55		G2-A22	0
G2-23		G2-A21	1	G2-56		na	-
G2-24		na	-	G2-57[f]	G2-A24	G2-A24	0
G2-25[c]	G2-A17	G2-A17	0	G2-58		G2-A23	0
G2-26[c]		G2-A17	0	G2-59		na	-
G2-27[c]		G2-A17	0	G2-60		na	-
G2-28[c]		G2-A17	0	G2-61	nd	na	-
G2-29[d]	G2-A18	G2-A18	0	G2-62		na	-
G2-30[d]		G2-A18	0	G2-63		na	-
G2-31[d]		G2-A18	0	G2-64		G2-A20	1
G2-32[d]		G2-A18	0				

Appendix 5-1C. Parent assignment for saplings from Malans. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching	Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
G3-01	G3-A1	na	-	G3-33	nd	na	-
G3-02		na	-	G3-34[f]		G3-A8	1
G3-03		na	-	G3-35[b]		G3-A7	1
G3-04		na	-	G3-36		na	-
G3-05[a]	G3-A14	G3-A14	0	G3-37	nd	na	-
G3-06[a]		G3-A14	0	G3-38		na	-
G3-07[a]		G3-A14	0	G3-39		na	-
G3-08[a]		G3-A14	0	G3-40		G3-A5	1
G3-09[b]	nd	G3-A7	0	G3-41	G3-A19	na	-
G3-10[c]		G3-A15	0	G3-42		na	-
G3-11[d]		G3-A12	2	G3-43		na	-
G3-12		na	-	G3-44		na	-
G3-13[e]	G3-A2	G3-A2	0	G3-45[c]	G3-A20	G3-A15	0
G3-14[e]		G3-A2	0	G3-46		na	-
G3-15		na	-	G3-47		G3-A18	0
G3-16		G3-A23	1	G3-48[c]		G3-A15	0
G3-17	G3-A16	na	-	G3-49	G3-A21	G3-A21	0
G3-18		na	-	G3-50		G3-A21	0
G3-19		na	-	G3-51		G3-A21	0
G3-20		G3-A19	1	G3-52		na	-
G3-21	nd	na	-	G3-53	G3-A23	na	-
G3-22		na	-	G3-54		G3-A4	2
G3-23		na	-	G3-55[c]		G3-A15	0
G3-24		na	-	G3-56		na	-
G3-25	G3-A17	na	-	G3-57	G3-A12	na	-
G3-26[c]		G3-A15	0	G3-58		G3-A4	1
G3-27		na	-	G3-59		G3-A17	1
G3-28		na	-	G3-60[d]		G3-A12	0
G3-29	nd	G3-A9	1	G3-61	G3-A24	na	-
G3-30		na	-	G3-62[f]		G3-A8	0
G3-31[c]		G3-A15	0	G3-63		G3-A24	0
G3-32		na	-	G3-64		na	-

Appendix 5-1D. Parent assignment for saplings from Mastrils. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching	Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
G4-01	nd	na	-	G4-33	G4-A11	na	-
G4-02		na	-	G4-34		na	-
G4-03[a]		G4-A1	2	G4-35		na	-
G4-04		na	-	G4-36		G4-A14	1
G4-05	nd	G4-A5	0	G4-37	nd	G4-A16	0
G4-07		na	-	G4-38		na	-
G4-08		na	-	G4-39		na	-
G4-09[b]	nd	G4-A24	1	G4-40[d]		G4-A25	1
G4-10		na	-	G4-41[a]	nd	G4-A1	1
G4-11		na	-	G4-42		G4-A13	0
G4-12		na	-	G4-43		na	-
G4-13	G4-A3	na	-	G4-44		na	-
G4-14		na	-	G4-45	G4-A12	na	-
G4-15		na	-	G4-46		na	-
G4-16		na	-	G4-47		na	-
G4-17[c]	G4-A5	G4-A18	0	G4-48		na	-
G4-18[c]		G4-A18	0	G4-49	nd	na	-
G4-19		na	-	G4-50		na	-
G4-20		na	-	G4-51[f]		G4-A15	1
G4-21	G4-A6	G4-A10	1	G4-53	G4-A14	G4-A3	1
G4-22		G4-A9	1	G4-54		na	-
G4-23		na	-	G4-55		na	-
G4-24		na	-	G4-56		G4-A21	1
G4-25[d]	G4-A7	G4-A25	1	G4-57[b]	nd	G4-A24	1
G4-26		na	-	G4-58		na	-
G4-27		na	-	G4-59[f]		G4-A15	0
G4-28[e]		G4-A12	1	G4-60		na	-
G4-29	G4-A8	G4-A17	1	G4-61	nd	na	-
G4-30		na	-	G4-62		na	-
G4-31		na	-	G4-63		na	-
G4-32[e]		G4-A12	1	G4-64		na	-

Appendix 5-1E. Parent assignment for saplings from Sargans. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching	Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
G5-01[a]	G5-A2	G5-A2	0	G5-33	nd	na	-
G5-02[a]		G5-A2	0	G5-34		na	-
G5-03[a]		G5-A2	0	G5-35		na	-
G5-04[a]		G5-A2	0	G5-36		na	-
G5-05[b]	G5-A14	G5-A17	1	G5-37	G5-A18	na	-
G5-06		na	-	G5-38		na	-
G5-07		G5-A14	0	G5-39		G5-A6	1
G5-08		na	-	G5-40		na	-
G5-09	G5-A16	G5-A12	1	G5-41	nd	na	-
G5-11		na	-	G5-42[c]		G5-A19	1
G5-12		G5-A15	0	G5-43		na	-
G5-13	G5-A3	na	-	G5-44		na	-
G5-14		na	-	G5-45	nd	na	-
G5-15[c]		G5-A19	0	G5-46[b]		G5-A17	1
G5-16		G5-A18	1	G5-47[c]		G5-A19	1
G5-17	nd	na	-	G5-48		na	-
G5-18		na	-	G5-49	G5-A19	na	-
G5-19		na	-	G5-50		G5-A16	0
G5-20		na	-	G5-51		na	-
G5-21	G5-A15	na	-	G5-52[a]		G5-A2	1
G5-22		G5-A24	1	G5-53	G5-A7	na	-
G5-23		G5-A25	1	G5-54		na	-
G5-24		na	-	G5-55		G5-A13	1
G5-25[d]	G5-A5	G5-A23	0	G5-56		G5-A21	1
G5-26[d]		G5-A23	0	G5-57	G5-A20	na	-
G5-27		na	-	G5-58		na	-
G5-28		na	-	G5-59		na	-
G5-29	nd	na	-	G5-60		na	-
G5-30		G5-A11	1	G5-61	G5-A9	na	-
G5-31		na	-	G5-62[b]		G5-A17	0
G5-32		na	-	G5-63		na	-
				G5-64		na	-

Appendix 5-1F. Parent assignment for saplings from Mels. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching	Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
G6-01		G6-A6	1	G6-33		na	-
G6-02	nd	na	-	G6-34[b]		G6-A8	0
G6-03[a]		G6-A17	1	G6-35	nd	na	-
G6-05[b]		G6-A8	0	G6-36		na	-
G6-06	nd	na	-	G6-37		na	-
G6-07		na	-	G6-38		na	-
G6-08		na	-	G6-39[e]	G6-A10	G6-A4	1
G6-09		na	-	G6-40[f]		G6-A2	2
G6-10	nd	na	-	G6-41[f]		G6-A2	1
G6-11		na	-	G6-43	G6-A11	G6-A11	0
G6-12		na	-	G6-44		na	-
G6-13		G6-A15	0	G6-46[d]		G6-A23	0
G6-14[a]	nd	G6-A17	1	G6-47	G6-A23	na	-
G6-15		na	-	G6-48[d]		G6-A23	0
G6-17[c]		G6-A13	1	G6-49[c]		G6-A13	0
G6-18[d]	nd	G6-A23	0	G6-50	G6-A12	na	-
G6-19		na	-	G6-51		na	-
G6-20		na	-	G6-52		na	-
G6-21[d]		G6-A23	1	G6-53[d]		G6-A23	1
G6-22	nd	na	-	G6-54	G6-A8	na	-
G6-23		na	-	G6-55		na	-
G6-24		na	-	G6-56		na	-
G6-25		G6-A9	2	G6-57		na	-
G6-26	G6-A9	na	-	G6-58	nd	na	-
G6-27[a]		G6-A17	1	G6-59[e]		G6-A4	1
G6-28		G6-A22	1	G6-60[a]		G6-A17	0
G6-29		na	-	G6-61		G6-A5	0
G6-30	nd	na	-	G6-62	nd	na	-
G6-31		G6-A18	0	G6-63		G6-A16	1
G6-32[d]		G6-A23	0	G6-64		G6-A25	1

Appendix 5-1G. Parent assignment for saplings from Ardon. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
W1-01[a]	W1-A4	W1-A6	1
W1-02		na	-
W1-03[b]		W1-A14	0
W1-04		W1-A20	1
W1-05	W1-A5	W1-A16	1
W1-06		na	-
W1-07		na	-
W1-08		W1-A3	1
W1-09	nd	na	-
W1-10		na	-
W1-11		na	-
W1-12		na	-
W1-13	W1-A19	na	-
W1-14		na	-
W1-15		na	-
W1-16		na	-
W1-17[c]	W1-A7	W1-A7	0
W1-18[c]		W1-A7	0
W1-19[c]		W1-A7	0
W1-20		na	-
W1-21	W1-A9	na	-
W1-22[d]		W1-A17	1
W1-23		na	-
W1-24		na	-
W1-25	W1-A7	na	-
W1-26		W1-A13	1
W1-27		W1-A11	1
W1-28		na	-
W1-29	W1-A12	na	-
W1-30		na	-
W1-31		na	-
W1-32[e]		W1-A15	0
W1-33[d]	nd	W1-A17	0
W1-34		W1-A23	0
W1-35[f]		W1-A25	0
W1-36[f]		W1-A25	0
W1-37	nd	na	-
W1-38		na	-
W1-40		na	-
W1-41	nd	na	-
W1-43a[d]		W1-A17	1
W1-43b		na	-
W1-44		na	-
W1-45	nd	na	-
W1-46		na	-
W1-47		na	-
W1-48[e]		W1-A15	0
W1-49	nd	na	-
W1-50		na	-
W1-51[c]		W1-A7	1
W1-52[b]		W1-A14	1
W1-53[c]	W1-A22	W1-A7	1
W1-54[c]		W1-A7	0
W1-55		na	-
W1-56		na	-
W1-57	W1-A3	W1-A22	1
W1-58[a]		W1-A6	1
W1-59		na	-
W1-60		na	-
W1-61	W1-A1	na	-
W1-62		na	-
W1-63		na	-
W1-64		na	-

Appendix 5-1H. Parent assignment for saplings from Chamoson. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching	Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
W2-01[a]	W2-A12	W2-A13	0	W2-33	nd	W2-A15	1
W2-02[b]		W2-A25	0	W2-34		na	-
W2-03		na	-	W2-35[i]		W2-A7	1
W2-04		na	-	W2-36		na	-
W2-05	W2-A13	W2-A23	1	W2-37	nd	na	-
W2-06[c]		W2-A12	0	W2-38		W2-A6	1
W2-07[c]		W2-A12	0	W2-39[g]		W2-A24	1
W2-08[b]		W2-A25	0	W2-40		na	-
W2-09	W2-A25	na	-	W2-41	W2-A17	na	-
W2-10		W2-A20	0	W2-42[j]		W2-A8	1
W2-11[c]		W2-A12	0	W2-43[j]		W2-A8	0
W2-12[d]		W2-A21	0	W2-44[j]		W2-A8	1
W2-13[e]	nd	W2-A16	1	W2-45[k]	nd	W2-A3	1
W2-14[f]		W2-A2	0	W2-46		W2-A11	1
W2-15		na	-	W2-47[d]		W2-A21	1
W2-16[f]		W2-A2	1	W2-48[k]		W2-A3	0
W2-17[a]	nd	W2-A13	0	W2-49[h]	W2-A6	W2-A22	0
W2-18[g]		W2-A24	0	W2-50		na	-
W2-19[h]		W2-A22	1	W2-51[h]		W2-A22	0
W2-20		na	-	W2-52[k]		W2-A3	1
W2-21	W2-A16	na	-	W2-53	W2-A11	W2-A17	1
W2-22[g]		W2-A24	0	W2-54		na	-
W2-23[e]		W2-A16	1	W2-55		na	-
W2-24[e]		W2-A16	1	W2-56		W2-A18	0
W2-25	W2-A3	na	-	W2-57	W2-A2	na	-
W2-26		na	-	W2-58		na	-
W2-27[c]		W2-A12	1	W2-59		na	-
W2-28		W2-A1	1	W2-60		na	-
W2-29	nd	na	-	W2-61	nd	na	-
W2-30		na	-	W2-62		na	-
W2-31		W2-A20	1	W2-63		W2-A5	0
W2-32[g]		W2-A24	2	W2-64[i]		W2-A7	1

Appendix 5-1I. Parent assignment for saplings from Saxon. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching	Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
W3-01	nd	na	-	W3-35	nd	na	-
W3-02		na	-	W3-36		W3-A13	1
W3-03	W3-A13	na	-	W3-37	nd	na	-
W3-04		na	-	W3-38[c]		W3-A18	0
W3-05	W3-A12	na	-	W3-39	nd	na	-
W3-06		na	-	W3-40		na	-
W3-07	nd	na	-	W3-41	nd	na	-
W3-08		na	-	W3-42[e]		W3-A1	1
W3-09[a]	W3-A9	W3-A9	0	W3-43	W3-A22	W3-A22	0
W3-10[a]		W3-A9	0	W3-44		W3-A17	1
W3-11[a]	nd	W3-A9	1	W3-45	W3-A23	na	-
W3-12		na	-	W3-46[f]		W3-A3	0
W3-13	W3-A21	na	-	W3-47[f]	nd	W3-A3	0
W3-14		na	-	W3-48		na	-
W3-15	W3-A18	na	-	W3-49[e]	nd	W3-A1	0
W3-16		na	-	W3-50		na	-
W3-17	nd	na	-	W3-53	nd	na	-
W3-18		W3-A11	0	W3-54[g]		W3-A23	0
W3-19	W3-A19	W3-A19	0	W3-55[h]	nd	W3-A24	0
W3-20		na	-	W3-56[h]		W3-A24	0
W3-21	W3-A20	na	-	W3-57[g]	nd	W3-A23	1
W3-22[a]		W3-A9	1	W3-58		na	-
W3-25[b]	W3-A4	W3-A4	0	W3-51	nd	na	-
W3-26[b]		W3-A4	0	W3-52		na	-
W3-27	W3-A1	na	-	W3-59		W3-A15	0
W3-28		W3-A14	1	W3-60		na	-
W3-29[c]	nd	W3-A18	0	W3-61	nd	na	-
W3-30		na	-	W3-62		W3-A16	1
W3-31[d]	W3-A7	W3-A7	0	W3-63	nd	na	-
W3-32[d]		W3-A7	0	W3-64		na	-
W3-33	W3-A8	na	-	W3-65	W3-A16	na	-
W3-34		na	-	W3-66	W3-A16	na	-

Appendix 5-1J. Parent assignment for saplings from Martigny. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching	Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
W4-01	nd	na	-	W4-33[c]	W4-A7	W4-A2	0
W4-02		na	-	W4-34[b]		W4-A12	1
W4-03	W4-A3	W4-A3	0	W4-35		W4-A7	0
W4-04		W4-A20	1	W4-36		na	-
W4-05		na	-	W4-37[c]	W4-A1	W4-A2	0
W4-06[a]		W4-A14	0	W4-38[c]		W4-A2	0
W4-07	W4-A10	na	-	W4-39		na	-
W4-08		na	-	W4-40		na	-
W4-09		na	-	W4-41	W4-A17	na	-
W4-10		na	-	W4-42		na	-
W4-11	nd	na	-	W4-43		na	-
W4-12		na	-	W4-44		na	-
W4-13		na	-	W4-45	nd	na	-
W4-14		na	-	W4-46		na	-
W4-15	W4-A12	W4-A4	0	W4-47		na	-
W4-16[b]		W4-A12	1	W4-48		na	-
W4-17[b]		W4-A12	0	W4-49	W4-A17	na	-
W4-18		W4-A11	2	W4-50		na	-
W4-19[a]	W4-A14	W4-A14	0	W4-51		na	-
W4-20[a]		W4-A14	0	W4-52		na	-
W4-21		na	-	W4-53	W4-A18	na	-
W4-22[c]		W4-A2	1	W4-54		W4-A19	0
W4-23	nd	na	-	W4-55		W4-A24	1
W4-24[d]		W4-A18	1	W4-56[d]		W4-A18	0
W4-25		na	-	W4-57	W4-A19	na	-
W4-26		W4-A17	0	W4-58		na	-
W4-27	nd	W4-A9	0	W4-59		W4-A1	0
W4-28		na	-	W4-60		na	-
W4-29	W4-A9	na	-	W4-61[c]	nd	W4-A2	1
W4-30		na	-	W4-62		W4-A8	1
W4-31		na	-	W4-63		na	-
W4-32		W4-A13	0	W4-64		na	-

Appendix 5-1K. Parent assignment for saplings from Collombey. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching	Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
W5-01[a]	W5-A1	W5-A1	0	W5-33[c]	W5-A10	W5-A10	0
W5-02[a]		W5-A1	0	W5-34[c]		W5-A10	0
W5-03[a]		W5-A1	0	W5-35[c]		W5-A10	0
W5-04[a]		W5-A1	0	W5-36[c]		W5-A10	0
W5-05[b]	W5-A2	W5-A2	0	W5-37[a]	W5-A11	W5-A1	1
W5-06[b]		W5-A2	0	W5-38		W5-A11	0
W5-07[b]		W5-A2	0	W5-39[e]		W5-A5	1
W5-08[b]		W5-A2	0	W5-40		na	-
W5-09[c]	W5-A3	W5-A10	0	W5-41[d]	W5-A12	W5-A17	1
W5-10		na	-	W5-42[b]		W5-A2	1
W5-11		na	-	W5-43		na	-
W5-12[c]		W5-A10	1	W5-44		na	-
W5-13	W5-A4	na	-	W5-45	nd	na	-
W5-14[d]		W5-A17	1	W5-46		na	-
W5-15[b]		W5-A2	1	W5-47		na	-
W5-16		W5-A15	1	W5-48		na	-
W5-17	nd	na	-	W5-49[d]	W5-A14	W5-A17	0
W5-18		na	-	W5-50		W5-A7	0
W5-19		W5-A12	1	W5-51		na	-
W5-20		na	-	W5-53[a]	W5-A18	W5-A1	0
W5-21[e]	W5-A5	W5-A5	0	W5-54		na	-
W5-22[e]		W5-A5	0	W5-55		W5-A18	0
W5-23[e]		W5-A5	0	W5-56		na	-
W5-24[e]		W5-A5	0	W5-57	W5-A19	na	-
W5-25[f]	W5-A6	W5-A6	0	W5-58		na	-
W5-26[f]		W5-A6	0	W5-59		na	-
W5-27[f]		W5-A6	0	W5-60		na	-
W5-28[f]		W5-A6	0	W5-61	W5-A20	na	-
W5-29[b]	W5-A9	W5-A2	1	W5-62		na	-
W5-30		na	-	W5-63		na	-
W5-31		na	-	W5-64		na	-
W5-32		W5-A16	0				

Appendix 5-1L. Parent assignment for saplings from Ollon. *Candidate parent* – adult tree that for geographical position could be the tree underneath which saplings where collected, *Assigned Parent* –Adult tree assigned as tentative parent by Cervus, *nd* – not determined, *na* - not assigned. Saplings with the same letter as superscript were assigned to the same tentative parent.

Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching	Sapling ID	Candidate parent ID	Assigned Parent ID	Number loci mismatching
W6-01	nd	na	-	W6-33[c]	nd	W6-A5	0
W6-02		W6-A9	1	W6-34		W6-A8	1
W6-03[a]		W6-A11	0	W6-35		na	-
W6-04[b]		W6-A16	0	W6-36		na	-
W6-05	nd	na	-	W6-37[e]	W6-A12	W6-A7	2
W6-06		na	-	W6-38		na	-
W6-07		na	-	W6-39		na	-
W6-08		na	-	W6-40		na	-
W6-09[b]	W6-A3	W6-A16	0	W6-41	W6-A10	W6-A18	1
W6-10[c]		W6-A5	1	W6-42		na	-
W6-11[a]		W6-A11	1	W6-43		na	-
W6-12		na	-	W6-44		na	-
W6-13	W6-A4	na	-	W6-45	W6-A11	na	-
W6-14		na	-	W6-46		na	-
W6-15		na	-	W6-47		na	-
W6-16[d]		W6-A4	1	W6-48		na	-
W6-17	W6-A5	na	-	W6-49	W6-A25	na	-
W6-18[d]		W6-A4	1	W6-50		na	-
W6-19		W6-A12	0	W6-51		na	-
W6-20		na	-	W6-52		na	-
W6-21[d]	nd	W6-A4	1	W6-53	W6-A24	na	-
W6-23[c]		W6-A5	1	W6-54[e]		W6-A7	1
W6-24		na	-	W6-55[d]		W6-A4	1
W6-25	W6-A6	na	-	W6-56		na	-
W6-26		na	-	W6-57	W6-A23	na	-
W6-27		na	-	W6-58[c]		W6-A5	0
W6-28		na	-	W6-59[a]		W6-A11	1
W6-29[a]	W6-A8	W6-A11	0	W6-60		na	-
W6-30[e]		W6-A7	0	W6-61	W6-A22	na	-
W6-31		na	-	W6-62		W6-A13	1
W6-32		na	-	W6-63		W6-A24	0
				W6-64		na	-

6. Synthesis

General results and discussion

The studied populations of *F. sylvatica* presented high genetic variability, as has been reported for other studies for both SSRs (Seifert 2012; Bontemps *et al.* 2013; Rajendra *et al.* 2014) and SNPs (Seifert *et al.* 2012; Müller *et al.* 2015a). The AMOVA analysis showed that 98-99% of the genetic variation is found within populations, while 1-2% is found among populations. Indeed, G''_{ST} values revealed low population differentiation, which was supported by the weak population structure revealed by STRUCTURE. Likewise, although there is a tendency for populations from the same valley to group together in NJ trees, the clusters were weakly supported by the bootstrap values, confirming the weak population differentiation and structure. These findings are in consensus with other studies also reporting low genetic differentiation for populations of beech (e.g., Buiteveld *et al.* 2007; Paffetti *et al.* 2012; Müller *et al.* 2015a). High genetic variability within populations and weak differentiation are explained by the outcrossing breeding system of this species and the high rates of pollen flow among populations (Oddou-Muratorio *et al.* 2011; Piotti *et al.* 2012). Interestingly, despite the low genetic differentiation, Chamoson, a population with low precipitation, seems to be more distinct, as revealed by both SSRs and SNPs in both saplings and adults. Some past management cannot be ruled out as a reason for this pattern.

A low percentage of loci in LD between pairs of SSR loci and SNPs loci was found. In forest trees, LD is low due to large population size and high outcrossing rates (Aitken *et al.* 2008). However, a higher percentage of pairs of SSR in LD was found in the saplings, which could be an effect of relatedness. Indeed, it was found that saplings collected underneath the same adult tree had higher pairwise relatedness coefficient. This is expected taking into account that in beech, seeds are primarily gravity dispersed and released around the mother tree (Millerón *et al.* 2013); thus, it is more likely that saplings collected at the same site had been produced by the same mother tree and consequently, more likely to be related.

Important climate related traits such as day of bud burst and drought tolerance were evaluated in the saplings. The day of bud burst was highly variable among populations, and it was influenced by the type of soil on which the saplings were growing: saplings on acidic soil flushed earlier. Drought tolerance was evaluated by, on the one hand, the indices F_V/F_M, PI_{abs}, PI_{tot}, which are based on chlorophyll fluorescence; and on the other hand by growth traits, specifically the increment in stem diameter (SG). Results show that the drought treatment affected negatively the performance of saplings, which was evident in the lower

values of PI_{abs}, PI_{tot} and SG. However, despite the negative effect of drought, populations from xeric environments performed better than populations from mesic environments. This is consistent with other studies reporting that beech trees from dry habitats are less affected by drought (Peuke *et al.* 2002; Fotelli *et al.* 2009; Weber *et al.* 2013; Thiel *et al.* 2014). Furthermore, it was found that the type of soil influences the response of the saplings. Saplings on acidic soil had earlier bud burst and also had lower SG, especially in the drought treatment. Soil characteristics influence not only the amount of water available for plants, but also the nutrients available for them (Geßler *et al.* 2007; Piedallu *et al.* 2013). Thus, soil characteristics affect the performance and sensitivity of plants under different environmental conditions.

Since populations from xeric environments were less sensitive to drought conditions, this might indicate that selective processes are acting on these populations. With the aim of identifying genetic adaptive variation in SNPs in candidate genes, three approaches were used: outlier analysis, phenotypic association analysis and environmental association analysis. These approaches are complementary because they exploit different sources of information, and although they generally yield different results concerning which loci should be considered under selection, it is considered that they are more likely to agree regarding true positives (De Mita *et al.* 2013; de Villemereuil *et al.* 2014) In this study, from the 76 SNPs genotyped, 6 were monomorphic (APX1_2, PhyB, 50_320, 52_1_249, 92_166, 110_1_111) and not included in the analyses. Then, from the remaining 70 SNPs, 40 SNPs (57.1%) were detected by any of the three approaches: 17 (24.3%) in the saplings and 35 in the adults (50%), and 13 SNPs (18.6%) were detected in both saplings and adults (Table 6-1). Loci under selection are likely to differ between different ages, because the environment can exert different selection pressures at different life stages (Petit & Hampe 2006), and also different sets of genes are involved in the same trait at different stages (Prunier *et al.* 2013). Besides, due to high competition and mortality, only a small fraction of seeds survive until the adult stage (Petit & Hampe 2006); thus, adult trees have passed different selection pressures through their life, and this could be evident in the higher number of SNPs likely to be under selection that were observed.

Nine SNPs (12.8%) were consistently identified by at least two approaches (Table 6-1); these SNPs are the most likely under selection in the studied populations. Most of the SNPs associated with environmental variables were found to be under positive selection according to the outlier analysis; in fact, for those SNPs, differences in allele frequencies were found in populations from contrasting environments.

Table 6-1 List of SNPs that show signatures of selection

Gene	SNP	SNP type	SAPLINGS			ADULTS	
			OA	PAA	EAA	OA	EAA
ALDH	ALDH_1[b]	Non-coding			T		T
	ALDH_2[b]	Non-Synonymous			T		T
	ALDH_4	Synonymous	Balancing				
IDH	IDH_1[a,b]	Synonymous		SG		Positive	T, P, H
	IDH_3	Non-coding					T, P, H
	IDH_4[a,b]	Synonymous		SG		Positive	T, P, H
APX	APX4_2	Non-Synonymous					T, P, H
ERD	ERD[a,b]	Non-coding	Positive				SR
Dhn	Dhn_1	Non-Synonymous				Balancing	
CP10	CP10_503	Synonymous				Balancing	
	CP10_1317	Non-coding					SR
	CP10_1428	Non-Synonymous					T
CysPro	CysPro_202[b]	Synonymous	Positive			Positive	
	CysPro_728	Non-coding				Balancing	
	CysPro_783	Non-coding					T
DAG	DAG_1059[a]	Synonymous				Positive	T
Histone3	His3C2_186	Non-coding					T
NAC	NAC_854[a,b]	Non-Synonymous			T, P, H, SR	Positive	
	NAC_962[b]	Synonymous	Positive			Positive	
	NAC_1300	Non-coding				Positive	
PP2C	PP2C_315	Non-Synonymous		F_V/F_M			
XTH	7_258[b]	Non-coding			SR		T, P, H
	7_520	Non-coding		F_V/F_M			
SDR	17_1081	Non-coding				Balancing	
KT2	39_282	Non-Synonymous					T, P, H
DREB	50_39	Non-Synonymous		SG, PI_{tot}			
	50_232[a,b]	Synonymous		SG		Balancing	
SAHH	52_1_235	Non-Synonymous					T, P, H
	52_1_368	Synonymous					T, P, H
GAPDH	68_277	Non-Synonymous					SR
CAT	91_2_57	Synonymous				Positive	
	91_2_141[a]	Synonymous				Positive	T, P, H
	91_2_231	Synonymous				Positive	
	91_2_448	Non-coding					T
	91_2_479	Non-coding				Positive	
ACC-oxidase	92_352[a]	Non-Synonymous				Positive	T, P, H
	92_630[a,b]	Non-coding			SR	Balancing	
sHsps	110_1_293	Synonymous		PI_{abs}			
	110_1_423	Non-Synonymous					T, P, H
PME	154_2_137[b]	Synonymous			T, P, H		SR

OA – Outlier Analysis, *PAA* – Phenotypic Association Analysis, *EAA* – Environmental Association Analysis, *SG* - Stem Growth, F_V/F_M - maximum quantum efficiency of PSII, PI_{abs} - performance index of PSII, PI_{tot} - total performance index of PSII, *T* - Temperature, *P* – Precipitation, *H* – Humidity, *SR* – Solar Radiation. [a] SNPs detected by at least two different methods; [b] SNPs detected in both saplings and adults.

Only two SNPs were identified by all the three approaches: IDH_1 and IDH_4 from the gene Isocitrate Dehydrogenase (IDH). The expression of this gene is induced by salt and drought stress (Liu *et al.* 2010; Leterrier *et al.* 2012), and high differentiation at this two SNPs has been detected in populations with different amount of precipitation in Germany (Seifert 2012). According to the outlier analysis, IDH_1 and IDH_4 are under positive selection and showed strong differences in allele frequencies in contrasting environments: allele C for IDH_1 and allele G for IDH_4 were more frequent in populations with drier conditions, higher maximum temperatures and higher number of summer and heat days. Additionally, for the phenotypic trait SG, those alleles conferred better performance to saplings under drought/calcareous conditions. This provides strong evidence for the involvement of genetic variability at gene isocitrate dehydrogenase in local adaptation of the studied populations of beech.

Non-synonymous SNPs cause a change in the aminoacid and thus, a change in the protein sequence, causing a direct effect on the phenotype. Therefore, they are considered the most likely target of natural selection. In this study, not only non-synonymous SNPs showed evidence of being under selection, but also synonymous SNPs as well as SNPs in non-coding regions were also identified (Table 6-1). Synonymous SNPs affect mRNA splicing, stability and translation kinetics (Pagani *et al.* 2005; Chamary *et al.* 2006; Komar 2007), and SNPs in non-coding regions are responsible for temporal and spatial regulation of gene expression (Barrett *et al.* 2012). Consequently, synonymous SNPs and non-coding SNPs could also represent adaptive genetic variation.

Genes with SNPs showing evidence of selection included a wide range of cellular functions, including oxidoreductases, hydrolases, oxidases, transferases, transporters, chaperones and transcription factors. This is expected since many traits in plants are polygenic, involving complex interactions among several genes (Ingvarsson & Street 2011). Besides, several SNPs at the same gene were identified by at least one of the approaches, and should not be disregarded for further investigation by using, for example, haplotypes, which can have a substantial advantage over single SNP analysis for the detection of adaptive genetic variation (Balding 2006).

Conclusions and perspectives

By using three different approaches, this study provided insights into the genetic adaptation potential of European beech to climate change as well as the genetic basis of climate change related traits.

The evaluation of phenotypic traits in the drought experiment carried out on the saplings showed that populations from xeric environments where less sensitive to drought. SNPs in candidate genes were found to be associated with the traits evaluated, and clearly some genotypes performed better than others. Additionally, SNPs associated with environmental variables were detected, and the allele frequencies at those SNPs differed between populations from contrasting environments. The outlier analysis indicated that most of those SNPs are under positive selection.

Among the SNPs detected, strong evidence indicated that the IDH_1 and IDH_4 SNPs from the Isocitrate Dehydrogenase gene are very likely subjected to selection, since alleles at this SNPs had higher frequency in populations in dry environments and conferred better performance to saplings in the drought treatment. However, since many traits in plants involve the interaction among several genes (Ingvarsson & Street 2011), SNPs in other genes should not be disregarded, as further research in different populations or using different analyses incorporating haplotypes or epistatic interactions (e.g., Csilléry *et al.* 2014) could prove their participation in adaptation to climate change conditions.

In this study, a candidate gene approach was used to investigate adaptive genetic variation in beech. The advantages of this approach is that it provides a direct link to particular genes, and thus, they are less demanding in terms of the number of markers required (Franks & Hoffmann 2012). Nevertheless, they are limited to the genes selected, overlooking important genetic variation at other genes that could be of relevance or unidentified genes. On the other hand, GWAS allow the identification of adaptive variation throughout the genome regardless of whether the function of the genes is previously known; however, they have lower statistical power (Korte & Farlow 2013). Besides, GWAS are only possible for species with reference genome available, something that is difficult to find for forest trees. This study demonstrated that candidate gene approaches are suitable for the study of genetic adaptive variability in forest trees, and provided insights into the genetic basis of adaptation to climate change in *F. sylvatica.*

Summary

European beech (*Fagus sylvatica*) is one of the most important forest tree species in Europe and could be affected by climate change. Climate change scenarios for Switzerland predict less precipitation, higher annual mean temperatures and more frequent droughts during summer that could affect beech survival. Additionally, increasing temperatures could promote earlier flushing in spring and later bud set in autumn, potentially increasing the risk of frost damage. Consequently, the genetic adaptation potential of European beech to climate change is of great interest.

The main objective of this study was to investigate the genetic basis of adaptation of European beech to climate change using a candidate gene approach. For this purpose, beech populations along precipitation gradients in Switzerland were selected. Samples of adult trees and saplings were collected, and the saplings were additionally subjected to a controlled drought stress experiment. Results of the drought stress experiment revealed that important climate related traits such as bud burst, chlorophyll fluorescence and stem growth of the saplings were negatively affected by drought conditions. However, saplings from xeric populations were less sensitive to drought conditions.

All individuals were genotyped for 13 microsatellite (SSR) markers and 76 SNPs in 24 climate-related candidate genes. Analyses of microsatellite and SNP markers demonstrated that the investigated populations have high genetic variability and low but significant population differentiation. Additionally, no significant differences in genetic variability were detected between saplings and adults. High genetic variability is important for a species to be able to adapt to environmental changes. Thus, to evaluate the adaptive genetic variability of European beech, three different approaches were used. First, outlier analysis revealed three outlier SSRs and three outlier SNPs that are potentially under positive or balancing selection. Second, a phenotypic association analysis revealed seven SNPs significantly associated with chlorophyll fluorescence and stem growth traits. The phenotypic variation explained by the significant associated SNPs was relatively high ($4.1 \leq R^2 \leq 13.4$) compared to other studies. Third, an environmental association analysis revealed 24 SNPs significantly associated with environmental variables such as precipitation, temperature and aridity. In total, 9 potentially adaptive SNPs in seven candidate genes were simultaneously identified by two or three of the approaches used; these SNPs might have the highest probability of being involved in adaptation to drought conditions. Two of the identified SNPs are non-synonymous, four are

synonymous and two are non-coding SNPs. Although synonymous and non-coding SNPs are traditionally thought to be neutral, they are involved in the regulation of gene expression and thus, can represent adaptive genetic variation.

Candidate gene approaches are limited by the selected genes and may ignore other genes that could be of relevance. However, for organisms for which there is no reference genome as is the case of European beech, candidate gene approaches are an excellent alternative for the study of genetic adaptive variation. The results obtained in this study contribute to a better understanding of the genetic adaptive potential of European beech to climate change and may improve the development of scientific guidelines for the sustainable management and the conservation of this important species.

Zusammenfassung

Die Rotbuche (*Fagus sylvatica*) ist eine der wichtigsten Laubbaumarten in Europa. Klimawandelmodelle sagen für die Schweiz geringere Niederschläge, eine höhere mittlere Jahrestemperatur und häufigere Trockenheitsereignisse während des Sommers voraus. Weiterhin können steigende Temperaturen zu einem früheren Blattaustrieb im Frühling und zu einer späteren Knospenbildung im Herbst führen, was zu einem erhöhten Risiko von Frostschäden führt. Diese veränderten Umweltfaktoren können das Wachstum und das Überleben der Rotbuche negativ beeinflussen. Daraus folgend, ist das genetische Adaptationspotential der Rotbuche in Bezug auf den Klimawandel von höchstem Interesse.

In dieser Studie wurde hauptsächlich die genetische Basis der Adaptation der Rotbuche an den Klimawandel mithilfe von Kandidatengenen untersucht. Dazu wurden Rotbuchenpopulationen entlang eines Niederschlagsgradienten in der Schweiz ausgesucht. In diesen Populationen wurde Blattmaterial von ausgewachsenen Bäume und von Keimlingen gesammelt. Die Keimlinge wurden zusätzlich für ein kontrolliertes Trockenstressexperiment genutzt. Die Ergebnisse des Trockenstressexperiments zeigten, dass wichtige klimarelevante Merkmale wie Knospensprung, Chlorophyllfluoreszenz und Stammwachstum der Keimlinge durch trockene Bedingungen negativ beeinflusst werden. Obwohl die Keimlinge von trockenen Populationen weniger sensitiv zu trockenen Bedingungen waren.

Alle Individuen wurden mit 13 Mikrosatellitenmarkern (SSR) und 76 SNPs in 24 klimaassoziierten Genen genotypisiert. Die Ergebnisse der Mikrosatelliten und SNP-Analysen zeigten eine hohe genetische Variabilität innerhalb, und eine geringe signifikante Differenzierung zwischen den untersuchten Populationen. Signifikante Unterschiede zwischen den ausgewachsenen Bäumen und Keimlingen konnte nicht gefunden werden. Hohe genetische Variabilität ist für eine Art wichtig um sich an Umweltveränderungen anpassen zu können. Daher wurden in dieser Studie drei Ansätze genutzt um die adaptive genetische Variabilität der Rotbuche zu untersuchen. Erstens, die Analyse von Ausreißern zeigte jeweils drei Ausreißer mit SSR-Markern und mit SNP-Markern, die potentiell unter positiver oder balancierter Selektion sind. Zweitens, bei einer Phänotyp-Assoziationsanalyse wurden sieben SNPs gefunden, die signifikant mit Chlorophyllfluoreszenz- und Stammzuwachs-Merkmalen assoziiert sind. Die phänotypische Variation, die durch die signifikant assoziierten SNPs erklärt wurde, war im Vergleich zu anderen Studien hoch ($4.1 \leq R^2 \leq 13.4$). Drittens, die Umweltassoziationsanalyse fand 24 SNPs, die signifikant mit Umweltfaktoren wie Niederschlag, Temperatur und Trockenheit assoziiert sind. Insgesamt

wurden neun potentiell adaptive SNPs in sieben Kandidatengenen gleichzeitig bei zwei oder drei der durchgeführten Ansätze identifiziert. Die Wahrscheinlichkeit bei der Adaptation an trockene Bedingungen involviert zu sein, ist für diese acht SNPs am höchsten. Zwei der neun identifizierten SNPs sind nicht-synonym, vier sind synonym und zwei sind nicht nicht-codierend. Obwohl synonyme und nicht-codierende SNPs traditionell als neutral angesehen werden, können sie in der Regulation von Genexpression involviert sein und daher adaptive genetische Variation repräsentieren.

Limitierend bei der Untersuchung von Kandidatengenen, ist die begrenzte Anzahl der selektierten und analysierten Gene und dabei das Ignorieren von anderen möglicherweise relevanter Gene. Jedoch bei Organismen, wie die Rotbuche, ohne verfügbares Referenzgenom, ist der Ansatz der Kandidatengenanalyse eine exzellente Alternative für die Untersuchung von genetischer adaptiver Variation. Die Ergebnisse dieser Studie können eine Hilfestellung sein, das genetische adaptive Potential der Rotbuche in Bezug auf den Klimawandel besser zu verstehen. Weiterhin können die erlangten Erkenntnisse zur Verbesserung von wissenschaftlichen Richtlinien zur nachhaltigen Bewirtschaftung und zum Schutz dieser wichtigen Baumart beitragen.

Bibliography

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

Alberto FJ, Derory J, Boury C *et al.* (2013) Imprints of Natural Selection Along Environmental Gradients in Phenology-Related Genes of Quercus petraea. *Genetics*, **195**, 495–512.

Allen CD, Macalady AK, Chenchouni H *et al.* (2010) A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management*, **259**, 660–684.

Alvarez N, Thiel-Egenter C, Tribsch A *et al.* (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632–640.

Anderson JT, Willis JH, Mitchell-Olds T (2011) Evolutionary genetics of plant adaptation. *Trends in Genetics : TIG*, **27**, 258–266.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.

Arend M, Gessler A, Schaub M (2016a) The influence of the soil on spring and autumn phenology in European beech. *Tree Physiology*, **36**, 78–85.

Arend M, Sever K, Pflug E, Gessler A, Schaub M (2016b) Seasonal photosynthetic response of European beech to severe summer drought: Limitation, recovery and post-drought stimulation. *Agricultural and Forest Meteorology*, **220**, 83–89.

Ashley MV (2010) Plant Parentage, Pollination, and Dispersal: How DNA Microsatellites Have Altered the Landscape. *Critical Reviews in Plant Sciences*, **29**, 148–161.

Asuka Y, Tani N, Tsumura Y, Tomaru N (2004) Development and characterization of microsatellite markers for Fagus crenata Blume. *Molecular Ecology Notes*, **4**, 101–103.

Badeck F-W, Bondeau A, Böttcher K *et al.* (2004) Responses of spring phenology to climate change. *New Phytologist*, **162**, 295–309.

Balding DJ (2006) A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*, **7**, 781–791.

Barrett LW, Fletcher S, Wilton SD (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences*, **69**, 3613–3634.

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, **12**, 767–780.

Baruck J, Nestroy O, Sartori G *et al.* (2016) Soil classification and mapping in the Alps: The current state and future challenges. *Geoderma*, **264, Part B**, 312–331.

Basler D, Körner C (2014) Photoperiod and temperature responses of bud swelling and bud burst in four temperate forest tree species. *Tree Physiology*, **00**, 1–12.

Beaumont MA (2005) Adaptation and speciation: what can Fst tell us? *Trends in Ecology & Evolution*, **20**, 435–440.

Beaumont MA, Nichols RA (1996) Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proceedings: Biological Sciences*, **263**, 1619–1626.

Belmonte J, Alarcón M, Avila A, Scialabba E, Pino D (2008) Long-range transport of beech (Fagus sylvatica L.) pollen to Catalonia (north-eastern Spain). *International Journal of Biometeorology*, **52**, 675–687.

Beniston M, Goyette S (2007) Changes in variability and persistence of climate in Switzerland: Exploring 20th century observations and 21st century simulations. *Global and Planetary Change*, **57**, 1–15.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Bhargava A, Fuentes FF (2010) Mutational Dynamics of Microsatellites. *Molecular Biotechnology*, **44**, 250–266.

Bilela S, Dounavi A, Fussi B *et al.* (2012) Natural regeneration of Fagus sylvatica L. adapts with maturation to warmer and drier microclimatic conditions. *Forest Ecology and Management*, **275**, 60–67.

Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution*, **18**, 503–511.

Bolte A, Czajkowski T, Kompa T (2007) The north-eastern distribution range of European beech—a review. *Forestry*, **80**, 413–429.

Bontemps A, Klein EK, Oddou-Muratorio S (2013) Shift of spatial patterns during early recruitment in Fagus sylvatica: Evidence from seed dispersal estimates based on genotypic data. *Forest Ecology and Management*, **305**, 67–76.

Bouriaud O, Bréda N, Moguédec GL, Nepveu G (2004) Modelling variability of wood density in beech as affected by ring age, radial growth and climate. *Trees*, **18**, 264–276.

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology*, **12**, 232.

Bradbury PJ, Zhang Z, Kroon DE *et al.* (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.

Buiteveld J, Vendramin GG, Leonardi S, Kamer K, Geburek T (2007) Genetic diversity and differentiation in European beech (Fagus sylvatica L.) stands varying in management history. *Forest Ecology and Management*, **247**, 98–106.

Caffarra A, Donnelly A (2011) The ecological significance of phenology in four different tree species: effects of light and temperature on bud burst. *International Journal of Biometeorology*, **55**, 711–721.

Cain ML, Milligan BG, Strand AE (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany*, **87**, 1217–1227.

Carnicer J, Coll M, Ninyerola M *et al.* (2011) Widespread crown condition decline, food web disruption, and amplified tree mortality with increased climate change-type drought. *Proceedings of the National Academy of Sciences*, **108**, 1474–1478.

CH2011 (2011) *Swiss Climate Change Scenarios CH2011*. C2SM, MeteoSwiss, ETH, NCCR Climate, and OcCC, Zurich, Switzerland.

CH2014-Impacts (2014) *Toward Quantitative Scenarios of Climate Change Impacts in Switzerland*. OCCR, FOEN, MeteoSwiss, C2SM, Agroscope, and ProClim, Bern, Switzerland.

Chakraborty S, Tiedemann AV, Teng PS (2000) Climate change: potential impact on plant diseases. *Environmental Pollution*, **108**, 317–326.

Chamary JV, Parmley JL, Hurst LD (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics*, **7**, 98–108.

Chen D, Ma X, Li C *et al.* (2014) A wheat aminocyclopropane-1-carboxylate oxidase gene, TaACO1, negatively regulates salinity stress in Arabidopsis thaliana. *Plant Cell Reports*, **33**, 1815–1827.

Chmura DJ, Anderson PD, Howe GT *et al.* (2011) Forest responses to climate change in the northwestern United States: Ecophysiological foundations for adaptive management. *Forest Ecology and Management*, **261**, 1121–1142.

Chybicki IJ, Trojankiewicz M, Oleksa A, Dzialuk A, Burczyk J (2009) Isolation-by-distance within naturally established populations of European beech (Fagus sylvatica). *Botany*, **87**, 791–798.

Cocozza C, de Miguel M, Pšidová E *et al.* (2016) Variation in Ecophysiological Traits and Drought Tolerance of Beech (Fagus sylvatica L.) Seedlings from Different Populations. *Frontiers in Plant Science*, **7**, 886.

Crookston NL, Rehfeldt GE, Dixon GE, Weiskittel AR (2010) Addressing climate change in the forest vegetation simulator to assess impacts on landscape forest dynamics. *Forest Ecology and Management*, **260**, 1198–1211.

Csilléry K, Lalagüe H, Vendramin GG *et al.* (2014) Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (Fagus sylvatica) populations. *Molecular Ecology*, **23**, 4696–4708.

Delhon C, Thiébault S (2005) The migration of beech (Fagus sylvatica L.) up the Rhone: the Mediterranean history of a "mountain" species. *Vegetation History and Archaeobotany*, **14**, 119–132.

Dell Inc (2015) *Dell Statistica (data analysis software system), version 12*. software.dell.com.

DeLucia EH, Nabity PD, Zavala JA, Berenbaum MR (2012) Climate Change: Resetting Plant-Insect Interactions. *Plant Physiology*, **160**, 1677–1685.

Devlin B, Roeder K (1999) Genomic Control for Association Studies. *Biometrics*, **55**, 997–1004.

Ding K, Zhang J, Zhou K, Shen Y, Zhang X (2005) htSNPer1.0: software for haplotype block partition and htSNPs selection. *BMC Bioinformatics*, **6**, 38.

Diniz-Filho JAF, Soares TN, Lima JS *et al.* (2013) Mantel test in population genetics. *Genetics and Molecular Biology*, **36**, 475–485.

Dittmar C, Zech W, Elling W (2003) Growth variations of Common beech (Fagus sylvatica L.) under different climatic and environmental conditions in Europe—a dendroecological study. *Forest Ecology and Management*, **173**, 63–78.

Dore MHI (2005) Climate change and changes in global precipitation patterns: What do we know? *Environment International*, **31**, 1167–1181.

Durand J, Bodénès C, Chancerel E *et al.* (2010) A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genomics*, **11**, 570.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Eckert AJ, Bower AD, González-Martínez SC *et al.* (2010a) Back to nature: ecological genomics of loblolly pine (Pinus taeda, Pinaceae). *Molecular Ecology*, **19**, 3789–3805.

Eckert AJ, Heerwaarden J van, Wegrzyn JL *et al.* (2010b) Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (Pinus taeda L., Pinaceae). *Genetics*, **185**, 969–982.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**, 435–445.

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity*, **99**, 125–132.

Emiliani G, Paffetti D, Vettori C, Giannini R (2004) Geographic distribution of genetic variability of Fagus sylvatica L. in southern Italian populations. *Forest Genetics (Slovak Republic)*, **11(3-4)**, 231–237.

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

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Fang J, Lechowicz MJ (2006) Climatic limits for the present distribution of beech (Fagus L.) species in the world. *Journal of Biogeography*, **33**, 1804–1819.

Foll M, 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.

Fotelli MN, Nahm M, Radoglou K *et al.* (2009) Seasonal and interannual ecophysiological responses of beech (Fagus sylvatica) at its south-eastern distribution limit in Europe. *Forest Ecology and Management*, **257**, 1157–1164.

Fotelli MN, Rennenberg H, Geßler A (2002) Effects of Drought on the Competitive Interference of an Early Successional Species (Rubus fruticosus) on Fagus sylvatica L. Seedlings: 15N Uptake and Partitioning, Responses of Amino Acids and other N Compounds. *Plant Biology*, **4**, 311–320.

Fotelli MN, Rienks M, Rennenberg H, Geßler A (2004) Climate and forest management affect 15N-uptake, N balance and biomass of European beech seedlings. *Trees*, **18**, 157–166.

Foulkes AS (2009) *Applied Statistical Genetics with R*. Springer New York, New York, NY.

François O, Martins H, Caye K, Schoville SD (2016) Controlling false discoveries in genome scans for selection. *Molecular Ecology*, **25**, 454–469.

Franks SJ, Hoffmann AA (2012) Genetics of Climate Change Adaptation. *Annual Review of Genetics*, **46**, 185–208.

Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, **6**, 925–929.

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Molecular Biology and Evolution*, **30**, 1687–1699.

Friedrichs DA, Trouet V, Büntgen U *et al.* (2009) Species-specific climate sensitivity of tree growth in Central-West Germany. *Trees*, **23**, 729–739.

Fu W, Akey JM (2013) Selection and adaptation in the human genome. *Annual Review of Genomics and Human Genetics*, **14**, 467–489.

Fu YH, Campioli M, Deckmyn G, Janssens IA (2012) The Impact of Winter and Spring Temperatures on Temperate Tree Budburst Dates: Results from an Experimental Climate Manipulation. *PLoS ONE*, **7**, e47324.

Führer E, Horvath L, Jagodics A, Machon A, Szabados I (2011) Application of a new aridity index in Hungarian forestry practice. *Idojárás*, **115**, 205–216.

Gailing O, Vornam B, Leinemann L, Finkeldey R (2009) Genetic and genomic approaches to assess adaptive genetic variation in plants: forest trees as a model. *Physiologia Plantarum*, **137**, 509–519.

Gallé A, Feller U (2007) Changes of photosynthetic traits in beech saplings (Fagus sylvatica) under severe drought stress and during recovery. *Physiologia Plantarum*, **131**, 412–421.

Gärtner S, Reif A, Xystrakis F *et al.* (2008) The drought tolerance limit of Fagus sylvatica forest on limestone in southwestern Germany. *Journal of Vegetation Science*, **19**, 757–768.

Gauzere J, Klein EK, Brendel O, Davi H, Oddou-Muratorio S (2016) Using partial genotyping to estimate the genetic and maternal determinants of adaptive traits in a progeny trial of Fagus sylvatica. *Tree Genetics & Genomes*, **12**, 115.

Geßler A, Keitel C, Kreuzwieser J *et al.* (2007) Potential risks for European beech (Fagus sylvatica L.) in a changing climate. *Trees*, **21**, 1–11.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463–471.

Goudet J (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *Journal of Heredity*, **86**, 485–486.

Grivet D, Sebastiani F, Alía R *et al.* (2011) Molecular Footprints of Local Adaptation in Two Mediterranean Conifers. *Molecular Biology and Evolution*, **28**, 101–116.

Haasl RJ, Payseur BA (2013) Microsatellites as Targets of Natural Selection. *Molecular Biology and Evolution*, **30**, 285–298.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Hao D, Chao M, Yin Z, Yu D (2012) Genome-wide association analysis detecting significant single nucleotide polymorphisms for chlorophyll and chlorophyll fluorescence parameters in soybean (Glycine max) landraces. *Euphytica*, **186**, 919–931.

Hartmann DL, Klein Tank AMG, Rusticucci M *et al.* (2013) Observations: Atmosphere and Surface. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker TF, Qin D, Plattner GK, et al.), pp. 159–254. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Hedrick PW (2001) Conservation genetics: where are we now? *Trends in Ecology & Evolution*, **16**, 629–636.

Hedrick PW (2005) *Genetics of Populations*. Jones & Bartlett Learning.

Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. *Nature Reviews Genetics*, **9**, 421–432.

Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. *Trends in Plant Science*, **15**, 675–683.

Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, **21**, 797–807.

Hong EP, Park JW (2012) Sample Size and Statistical Power Calculation in Genetic Association Studies. *Genomics & Informatics*, **10**, 117–122.

Houston Durrant T, de Rigo D, Caudullo G (2016) Fagus sylvatica and other beeches in Europe: distribution, habitat, usage and threats. In: *European Atlas of Forest Tree Species* (eds San-Miguel-Ayanz J, de Rigo D, Caudullo G, Houston Durrant T, Mauri A), p. e12b90+. Publ. Off. EU, Luxembourg.

Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. *New Phytologist*, **189**, 909–922.

IPCC (2014) *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (RK Pachauri, LA Meyer, Eds,). IPCC, Geneva, Switzerland.

Jahn G (1991) Temperate deciduous forests. In: *Ecosystems of the world* (eds Röhrig E, Ulrich B), pp. 377–502. Elsevier, Amsterdam, The Netherlands.

Jazbec A, Šegotić K, Ivanković M, Marjanović H, Perić S (2007) Ranking of European beech provenances in Croatia using statistical analysis and analytical hierarchy process. *Forestry*, **80**, 151–162.

Jensen TS (1985) Seed-Seed Predator Interactions of European Beech, Fagus silvatica and Forest Rodents, Clethrionomys glareolus and Apodemus flavicollis. *Oikos*, **44**, 149–156.

Johnson JS, Gaddis KD, Cairns DM, Krutovsky KV (2017) Seed dispersal at alpine treeline: an assessment of seed movement within the alpine treeline ecotone. *Ecosphere*, **8**, e01649.

Jones AG, Small CM, Paczolt KA, Ratterman NL (2010) A practical guide to methods of parentage analysis. *Molecular Ecology Resources*, **10**, 6–30.

Jump AS, Hunt JM, Martínez-Izquierdo JA, Peñuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in Fagus sylvatica. *Molecular Ecology*, **15**, 3469–3480.

Jump AS, Peñuelas J (2007) Extensive spatial genetic structure revealed by AFLP but not SSR molecular markers in the wind-pollinated tree, Fagus sylvatica. *Molecular Ecology*, **16**, 925–936.

Kalinowski ST (2005) hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Kirk H, Freeland JR (2011) Applications and Implications of Neutral versus Non-neutral Markers in Molecular Ecology. *International Journal of Molecular Sciences*, **12**, 3966–3988.

Köcher P, Gebauer T, Horna V, Leuschner C (2009) Leaf water status and stem xylem flux in relation to soil drought in five temperate broad-leaved tree species with contrasting water use strategies. *Annals of Forest Science*, **66**, 1, 101.

Komar AA (2007) Genetics. SNPs, silent but not invisible. *Science (New York, N.Y.)*, **315**, 466–467.

Konijnendijk N, Shikano T, Daneels D, Volckaert FAM, Raeymaekers JAM (2015) Signatures of selection in the three-spined stickleback along a small-scale brackish water – freshwater transition zone. *Ecology and Evolution*, **5**, 4174–4186.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**, 1179–1191.

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, **9**, 29.

Kovats RS, Valentini R, Bouwer LM *et al.* (2014) Europe. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*

(eds Barros V., Field CB, Dokken DJ, et al.), pp. 1267–1326. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kraj W, Sztorc A (2009) Genetic structure and variability of phenological forms in the European beech (Fagus sylvatica L.). *Annals of Forest Science*, **66**, 203–203.

Kramer K, Degen B, Buschbom J *et al.* (2010) Modelling exploration of the future of European beech (Fagus sylvatica L.) under climate change—Range, abundance, genetic diversity and adaptive response. *Forest Ecology and Management*, **259**, 2213–2222.

Kremer A, Ronce O, Robledo-Arnuncio JJ *et al.* (2012) Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters*, **15**, 378–392.

Krutovsky KV, Clair JBS, Saich R, Hipkins VD, Neale DB (2009) Estimation of population structure in coastal Douglas-fir [Pseudotsuga menziesii (Mirb.) Franco var. menziesii] using allozyme and microsatellite markers. *Tree Genetics & Genomes*, **5**, 641–658.

Kunstler G, Thuiller W, Curt T *et al.* (2007) Fagus sylvatica L. recruitment across a fragmented Mediterranean Landscape, importance of long distance effective dispersal, abiotic conditions and biotic interactions. *Diversity and Distributions*, **13**, 799–807.

Kuster TM, Arend M, Bleuler P, Günthardt-Goerg MS, Schulin R (2013) Water regime and growth of young oak stands subjected to air-warming and drought on two different forest soils in a model ecosystem experiment. *Plant Biology*, **15**, 138–147.

Lalagüe H, Csilléry K, Oddou-Muratorio S *et al.* (2014) Nucleotide diversity and linkage disequilibrium at 58 stress response and phenology candidate genes in a European beech (Fagus sylvatica L.) population from southeastern France. *Tree Genetics & Genomes*, **10**, 15–26.

Langella, O (1999) Populations 1.2.31.

Lebourgeois F, Bréda N, Ulrich E, Granier A (2005) Climate-tree-growth relationships of European beech (Fagus sylvatica L.) in the French Permanent Plot Network (RENECOFOR). *Trees*, **19**, 385–401.

Lefèvre S, Wagner S, Petit RJ, De LAFONTAINE G (2012) Multiplexed microsatellite markers for genetic studies of beech. *Molecular Ecology Resources*, **12**, 484–491.

Lehner B, Döll P, Alcamo J, Henrichs T, Kaspar F (2006) Estimating the Impact of Global Change on Flood and Drought Risks in Europe: A Continental, Integrated Analysis. *Climatic Change*, **75**, 273–299.

Leterrier M, Barroso JB, Valderrama R, Palma JM, Corpas FJ (2012) NADP-Dependent Isocitrate Dehydrogenase from Arabidopsis Roots Contributes in the Mechanism of Defence against the Nitro-Oxidative Stress Induced by Salinity. *The Scientific World Journal*, **2012**, e694740.

Leuschner C, Backes K, Hertel D *et al.* (2001) Drought responses at leaf, stem and fine root levels of competitive Fagus sylvatica L. and Quercus petraea (Matt.) Liebl. trees in dry and wet years. *Forest Ecology and Management*, **149**, 33–46.

Lewontin RC, Krakauer J (1973) Distribution of Gene Frequency as a Test of the Theory of the Selective Neutrality of Polymorphisms. *Genetics*, **74**, 175–195.

Li J, Li H, Jakobsson M *et al.* (2012) Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Molecular Ecology*, **21**, 28–44.

Liu Y, Shi Y, Song Y, Wang T, Li Y (2010) Characterization of a Stress-induced NADP-isocitrate Dehydrogenase Gene in Maize Confers Salt Tolerance in Arabidopsis. *Journal of Plant Biology*, **53**, 107–112.

Magri D (2008) Patterns of post-glacial spread and the extent of glacial refugia of European beech (Fagus sylvatica). *Journal of Biogeography*, **35**, 450–463.

Magri D, Vendramin GG, Comps B *et al.* (2006) A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist*, **171**, 199–221.

Maliva R, Missimer T (2012) Aridity and Drought. In: *Arid Lands Water Evaluation and Management* Environmental Science and Engineering., pp. 21–39. Springer Berlin Heidelberg.

Manel S, Gugerli F, Thuiller W *et al.* (2012) Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology*, **21**, 3729–3738.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. *Molecular Ecology*, **19**, 3824–3835.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639–655.

Maxwell K, Johnson GN (2000) Chlorophyll fluorescence—a practical guide. *Journal of Experimental Botany*, **51**, 659–668.

McCune B, Keon D (2002) Equations for potential annual direct incident radiation and heat load. *Journal of Vegetation Science*, **13**, 603–606.

Meirmans PG, Hedrick PW (2011) Assessing population structure: FST and related measures. *Molecular Ecology Resources*, **11**, 5–18.

Milad M, Schaich H, Bürgi M, Konold W (2011) Climate change and nature conservation in Central European forests: A review of consequences, concepts and challenges. *Forest Ecology and Management*, **261**, 829–843.

Millerón M, López de Heredia U, Lorenzo Z *et al.* (2013) Assessment of spatial discordance of primary and effective seed dispersal of European beech (Fagus sylvatica L.) by ecological and genetic methods. *Molecular Ecology*, **22**, 1531–1545.

De Mita S, Thuillet A-C, Gay L *et al.* (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, **22**, 1383–1399.

Morin PA, Luikart G, Wayne RK, group the S workshop (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.

Müller M (2014) A candidate gene-based association study to investigate potentially adaptive genetic variation in European beech (Fagus sylvatica L.). Dissertation Thesis. Georg-August-University Göttingen.

Müller M, Seifert S, Finkeldey R (2015a) A candidate gene-based association study reveals SNPs significantly associated with bud burst in European beech (Fagus sylvatica L.). *Tree Genetics & Genomes*, **11**, 1–13.

Müller M, Seifert S, Finkeldey R (2015b) Identification of SNPs in candidate genes potentially involved in bud burst in European beech (Fagus sylvatica L.). *Silvae Genetica*, **64**, 1–20.

Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology*, **17**, 3599–3613.

Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, **11**, 184–194.

Nei M (1972) Genetic Distance between Populations. *The American Naturalist*, **106**, 283–292.

Nielsen R (2005) Molecular Signatures of Natural Selection. *Annual Review of Genetics*, **39**, 197–218.

Nilsson SG (1985) Ecological and Evolutionary Interactions between Reproduction of Beech Fagus silvatica and Seed Eating Animals. *Oikos*, **44**, 157–164.

Noguero M, Atif RM, Ochatt S, Thompson RD (2013) The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. *Plant Science*, **209**, 32–45.

Oddou-Muratorio S, Klein EK, Vendramin GG, Fady B (2011) Spatial vs. temporal effects on demographic and genetic structures: the roles of dispersal, masting and differential mortality on patterns of recruitment in Fagus sylvatica. *Molecular Ecology*, **20**, 1997–2010.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Packham JR, Thomas PA, Atkinson MD, Degen T (2012) Biological Flora of the British Isles: Fagus sylvatica. *Journal of Ecology*, **100**, 1557–1608.

Paffetti D, Travaglini D, Buonamici A *et al.* (2012) The influence of forest management on beech (Fagus sylvatica L.) stand structure and genetic diversity. *Forest Ecology and Management*, **284**, 34–44.

Pagani F, Raponi M, Baralle FE (2005) Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6368–6372.

Pastorelli R, Smulders MJM, VAN'T Westende WPC *et al.* (2003) Characterization of microsatellite markers in Fagus sylvatica L. and Fagus orientalis Lipsky. *Molecular Ecology Notes*, **3**, 76–78.

Peakall R, Smouse PE (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, **28**, 2537–2539.

Pearson T., Manolio T. (2008) How to interpret a genome-wide association study. *JAMA*, **299**, 1335–1344.

Perea R, Miguel AS, Gil L (2011) Flying vs. climbing: Factors controlling arboreal seed removal in oak–beech forests. *Forest Ecology and Management*, **262**, 1251–1257.

Petit RJ, Hampe A (2006) Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 187–214.

Peuke AD, Schraml C, Hartung W, Rennenberg H (2002) Identification of drought-sensitive beech ecotypes by physiological parameters. *New Phytologist*, **154**, 373–387.

Piedallu C, Gégout J-C, Perez V, Lebourgeois F (2013) Soil water balance performs better than climatic water variables in tree species distribution modelling. *Global Ecology and Biogeography*, **22**, 470–482.

Piotti A, Leonardi S, Buiteveld J *et al.* (2012) Comparison of pollen gene flow among four European beech (Fagus sylvatica L.) populations characterized by different management regimes. *Heredity*, **108**, 322–331.

Piotti A, Leonardi S, Heuertz M *et al.* (2013) Within-Population Genetic Structure in Beech (Fagus sylvatica L.) Stands Characterized by Different Disturbance Histories: Does Forest Management Simplify Population Substructure? *PLoS ONE*, **8**, e73391.

Piovesan G, Biondi F, Filippo AD, Alessandrini A, Maugeri M (2008) Drought-driven growth reduction in old beech (Fagus sylvatica L.) forests of the central Apennines, Italy. *Global Change Biology*, **14**, 1265–1281.

Pluess AR, Frank A, Heiri C *et al.* (2016) Genome–environment association study suggests local adaptation to climate at the regional scale in Fagus sylvatica. *New Phytologist*, **210**, 589–601.

Pluess AR, Määttänen K (2013) Characterization of eighteen novel microsatellite markers and multiplex PCR protocol for Fagus sylvatica. *Conservation Genetics Resources*, **5**, 311–314.

Pluess AR, Weber P (2012) Drought-Adaptation Potential in Fagus sylvatica: Linking Moisture Availability with Genetic Diversity and Dendrochronology. *PLoS ONE*, **7**, e33636.

Poncet BN, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of Arabis alpina. *Molecular Ecology*, **19**, 2896–2907.

Porth I, Klapšte J, Skyba O *et al.* (2013) Genome-wide association mapping for wood characteristics in Populus identifies an array of candidate single nucleotide polymorphisms. *New Phytologist*, **200**, 710–726.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, **155**, 945–959.

Priwitzer T, Kurjak D, Kmeť J, Sitková Z, Leštianska A (2014) Photosynthetic response of European beech to atmospheric and soil drought. *Forestry Journal*, **60**, 32–38.

Prunier J, Laroche J, Beaulieu J, 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*, **20**, 1702–1716.

Prunier J, Pelgas B, Gagnon F *et al.* (2013) The genomic architecture and association genetics of adaptive characters using a candidate SNP approach in boreal black spruce. *BMC Genomics*, **14**, 368.

Putman AI, Carbone I (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, **4**, 4399–4428.

Queller DC, Goodnight KF (1989) Estimating Relatedness Using Genetic Markers. *Evolution*, **43**, 258–275.

R Core Team (2016) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rajendra KC, Seifert S, Prinz K, Gailing O, Finkeldey R (2014) Subtle human impacts on neutral genetic diversity and spatial patterns of genetic variation in European beech (Fagus sylvatica). *Forest Ecology and Management*, **319**, 138–149.

Rambaut, A (2014) FigTree.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348–4370.

Robson TM, Rodríguez-Calcerrada J, Sánchez-Gómez D, Aranda I (2009) Summer drought impedes beech seedling performance more in a sub-Mediterranean forest understory than in small gaps. *Tree Physiology*, **29**, 249–259.

Rose L, Leuschner C, Köckemann B, Buschmann H (2009) Are marginal beech (Fagus sylvatica L.) provenances a source for drought tolerant ecotypes? *European Journal of Forest Research*, **128**, 335–343.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Russello MA, Kirk SL, Frazer KK, Askey PJ (2012) Detection of outlier loci and their utility for fisheries management. *Evolutionary Applications*, **5**, 39–52.

Sander T, König S, Rothe GM, Janssen A, Weisgerber H (2000) Genetic variation of European beech (Fagus sylvatica L.) along an altitudinal transect at Mount Vogelsberg in Hesse, Germany. *Molecular Ecology*, **9**, 1349–1361.

Scharnweber T, Manthey M, Criegee C *et al.* (2011) Drought matters – Declining precipitation influences growth of Fagus sylvatica L. and Quercus robur L. in north-eastern Germany. *Forest Ecology and Management*, **262**, 947–961.

Schoville SD, Bonin A, François O *et al.* (2012) Adaptive Genetic Variation on the Landscape: Methods and Cases. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 23–43.

Schröder W, Schmidt G, Hasenclever J (2006) Geostatistical Analysis of data on AIR Temperature and Plant Phenology From Baden-Württemberg (GERMANY) as a Basis for Regional Scaled Models of Climate Change. *Environmental Monitoring and Assessment*, **120**, 27–43.

Seifert S (2012) Variation of candidate genes related to climate change in European beech (*Fagus sylvatica* L.). Dissertation Thesis. Georg-August-University Göttingen.

Seifert S, Vornam B, Finkeldey R (2012) A set of 17 single nucleotide polymorphism (SNP) markers for European beech (Fagus sylvatica L.). *Conservation Genetics Resources*, **4**, 1045–1047.

Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. *Journal of Experimental Botany*, **58**, 221–227.

Sillanpää MJ (2011) Overview of techniques to account for confounding due to population stratification and cryptic relatedness in genomic data association analyses. *Heredity*, **106**, 511–519.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, **82**, 561–573.

Sofo A, Scopa A, Nuzzaci M, Vitti A (2015) Ascorbate Peroxidase and Catalase Activities and Their Genetic Regulation in Plants Subjected to Drought and Salinity Stresses. *International Journal of Molecular Sciences*, **16**, 13561–13578.

Sork VL, Davis FW, Westfall R *et al.* (2010) Gene movement and genetic association with regional climate gradients in California valley oak (Quercus lobata Née) in the face of climate change. *Molecular Ecology*, **19**, 3806–3823.

Stephan W (2016) Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology*, **25**, 79–88.

Stone AC, Gehring CA, Whitham TG (2010) Drought negatively affects communities on a foundation tree: growth rings predict diversity. *Oecologia*, **164**, 751–761.

Swaty RL, Deckert RJ, Whitham TG, Gehring CA (2004) Ectomycorrhizal Abundance and Community Composition Shifts with Drought: Predictions from Tree Rings. *Ecology*, **85**, 1072–1084.

Thiel D, Kreyling J, Backhaus S *et al.* (2014) Different reactions of central and marginal provenances of Fagus sylvatica. *European Journal of Forest Research*, **133**, 247–260.

Thornthwaite CW (1948) An Approach toward a Rational Classification of Climate. *Geographical Review*, **38**, 55–94.

Trenberth K (2011) Changes in precipitation with climate change. *Climate Research*, **47**, 123–138.

Trotter RT, Cobb NS, Whitham TG (2008) Arthropod community diversity and trophic structure. *Ecological Entomology*, **33**, 1–11.

Tsumura Y, Uchiyama K, Moriguchi Y *et al.* (2014) Genetic Differentiation and Evolutionary Adaptation in Cryptomeria japonica. *G3: Genes|Genomes|Genetics*, **4**, 2389–2402.

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology*, **23**, 48–55.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, **34**, 275.

De Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology*, **23**, 2006–2019.

Vitasse Y, Delzon S, Bresson CC, Michalet R, Kremer A (2009) Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. *Canadian Journal of Forest Research*, **39**, 1259–1269.

Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting Natural Selection in Genomic Data. *Annual Review of Genetics*, **47**, 97–120.

Vornam B, Decarli N, Gailing O (2004) Spatial Distribution of Genetic Variation in a Natural Beech Stand (Fagus sylvaticaL.) Based on Microsatellite Markers. *Conservation Genetics*, **5**, 561–570.

Wang KS (2004) Gene Flow in European Beech (Fagus sylvatica L.). *Genetica*, **122**, 105–113.

Weber P, Bugmann H, Pluess AR, Walthert L, Rigling A (2013) Drought response and changing mean sensitivity of European beech close to the dry distribution limit. *Trees*, **27**, 171–181.

Weber P, Pluess, Andrea, Mühlethaler, Urs (2010) Genetic resources of beech in Switzerland. COST Action E 52 report in 'Status of the Genetic Resources of European Beech'. *Communicationes Instituti Forestalis Bohemicae*, **25**, 248–255.

Weir BS, Anderson AD, Hepler AB (2006) Genetic relatedness analysis: modern data and new challenges. *Nature Reviews Genetics*, **7**, 771–780.

Whitlock MC (2005) Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of Evolutionary Biology*, **18**, 1368–1373.

Wielgolaski FE (2001) Phenological modifications in plants by various edaphic factors. *International Journal of Biometeorology*, **45**, 196–202.

Von Wuehlisch G (2008) EUFORGEN Technical guidelines for genetic conservation and use for European beech (Fagus sylvatica). *Biodiversity International*, 6.

Zhao M, Running SW (2010) Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009. *Science*, **329**, 940–943.

Declaration of Honor

I hereby declare that my doctoral thesis entitled **“Genetic analysis of European beech populations across precipitation gradients: understanding the adaptive potential to climate change”** has been written independently and with no other sources and aids than quoted. I have indicated the parts that were performed by project collaborators. I further declare that this work has never been submitted in any form as part of other dissertation procedures.

Laura Carolina Cuervo Alarcon

Göttingen, 2017

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name	**Cuervo Alarcon, Laura Carolina**
Date of Birth	8 December 1985
Nationality	Colombian
E-mail	lcuervo@forst.uni-goettingen.de lccuervoa@gmail.com

EDUCATION

Oct. 2013-March 2017	**PhD student** in the section Forest Genetics and Forest Tree Breeding, Georg-August-University of Göttingen, Germany
2008-2010	**Master in Sciences-Biology** National University of Colombia, Bogotá
2003-2008	**Biology** National University of Colombia, Bogotá

PUBLICATIONS

Cuervo, L., Burbano, C. 2012. Caracterización Genética de la población *ex situ* de *Crocodylus intermedius* presente en Colombia. Rev. Acad. Colomb. Cienc. **36** (140): 373-383.

Palacios, C., **Cuervo, L.**, Cadavid, L. 2011. Evolutionary patterns of killer cell Ig-like receptor genes in Old World monkeys. Gene. 474, 39-51.

CONFERENCES

- **Cuervo-Alarcon, L.,** Arend, M., Müller, M., Finkeldey, R., Krutovsky, K.V. Adaptation potential of European beech to climate change: linking genetic variability with drought related traits. 2nd Annual Meeting in Conservation Genetics. October 5-6 2016, Göttingen, Germany (**Oral presentation**).

- **Cuervo-Alarcon, L.,** Arend, M., Müller, M., Finkeldey, R., Krutovsky, K.V. Genetic analysis of European beech populations across a precipitation gradient: understanding the adaptive potential to climate change. Annual Meeting of the Ecological Society of Germany, Austria and Switzerland. 31 August - 3 September 2015, Göttingen, Germany (**Oral presentation**).

- **Cuervo-Alarcon, L.,** Arend, M., Müller, M., Finkeldey, R., Krutovsky, K.V. Genetic analysis of European beech populations across a precipitation gradient using microsatellite markers and SNPs in candidate genes. Conference Global Change and the Evolutionary Potential of Forest Trees. August 24-25 2015, Iceland (**Oral presentation**).

- **Cuervo, L.,** Burbano, C**.** Genetic characterization of *Crocodylus intermedius* Colombian *ex situ* population by microsatellites. III Colombian Congress of Zoology. November-2010, Medellín-Colombia **(Oral presentation)**.

- **Cuervo, L.,** Burbano, C. Genetic characterization of *Crocodylus intermedius* (Graves, 1819) Colombian *ex situ* population. I International Congress of Orinoco´s basin biodiversity. November-2009, Villavicencio-Colombia **(Oral presentation).**

- Palacios, D.C., **Cuervo, L.,** Cadavid, L.F. NK cell immunoglobulin-like receptors evolution in old world monkeys. X Colombian Congress of Human Genetics. September-2009, Villa de Leyva-Colombia. **(Poster Presentation).**

- Victoria, D.M., Herrera, M., **Cuervo, L.,** Spinel, C. Ultrastructural Study of Acini of Submandibular Gland of Mouse in Culture. 10th Inter-American Congress of Electron Microscopy CIASEM 2009. October-2009, Argentina. **(Poster Presentation).**

- Victoria, D.M., **Cuervo, L.,** Spinel, C.M. Morphological analysis of Acini Closed in the long Term of Submandibular gland mouse. American Society for Cell Biology`s 47th Annual Meeting. December-2007, Washington D.C-EEUU. **(Poster Presentation).**

- **Cuervo-Alarcón, L.C.,** Low-Calle A.M. Initial patterns of primary succession on La Playa river rocks, Chingaza National Naural Park. I National Congress V Scientific Meeting of Biology Students. October-2007, Bogotá-Colombia. **(Oral presentation).**

LANGUAGES

Spanish	Native speaker
English	Fluent