

**Charles University**  
**Department of Botany, Faculty of Science**

Study programme: Evolutionary biology



**Bc. Alžběta Poštulková**

The role of the *FAR5* gene in alpine adaptation  
of *Arabidopsis arenosa*

Úloha genu *FAR5* v horské adaptaci *Arabidopsis arenosa*

Diploma thesis

Supervisor: Mgr. Magdalena Bohutínská, Ph.D.

Prague, 2024



***Declaration***

*I declare that I prepared the final thesis independently and that I have properly acknowledged and cited all the information sources and literature used. This work, or a substantial part of it, has not been submitted for the award of another or the same academic degree.*

*In Prague, 30. 4. 2024*

*Alžběta Poštulková*



# Acknowledgements

I would like to take this opportunity to express my infinite thanks to my wonderful supervisor, Majda Bohutínská, for her ideas, corrections and especially her patience during the process of creating this thesis. Her endless enthusiasm is an endless inspiration for me. I hope that I will make her proud in my coming years in Majdalab.

Further, I thank Erwann Arc for his help with preparation, management and evaluation of the transplant experiment in Innsbruck. This project would never be possible without him. With this I would also like to thank Ilse Kranner for letting me be a part of her lab for two months. It was a wonderful experience.

Many thanks to Filip Kolář for sharing his resources and knowledge during my project and to all the members of his Plant ecological genomics team for helping me with my megalomaniacal experimental plans. Big thanks to Gábi Šrámková and Verča Vlčková for their help with optimizing and starting my experiment and for their guidance throughout the practical part of my thesis.

Great thanks belong also to Frédéric Domergue for sharing his invaluable knowledge about *FAR5* with me.

I would also like to thank my family for their support, help and curiosity about my progress. Special thanks to my mom for instilling my love for nature and to my sister for always being there for me and my academic problems.

The warmest thanks I can give belong to my beloved Patrik for his constant support. He was great throughout the whole time, always by my side, caring for me and feeding me in the course of finishing my thesis. And also, for saving my experiment many times by willing to water, cross and even genotype the plants. He is simply the best!

Finally, I would like to thank Amy Rumpíková and Martin Buchta for being interested in my project and helping me with some difficult tasks throughout the project.

*This project was financially supported by Charles University Grant Agency (GAUK 219223). Part of the project was conducted during my stay at the University of Innsbruck and supported by The Mobility Fund and The Endowment Fund of the Faculty of Science of Charles University.*



# Abstract

The repeated adaptation presents a unique opportunity to study the mechanisms of evolution in natural replicates. Repeated adaptation of *Arabidopsis arenosa*, a wild relative of the model organism *Arabidopsis thaliana*, to the alpine environment was previously studied. Genomic analysis by selection scans revealed a set of candidate alleles. Here, I present a functional follow-up study of a candidate alpine-adaptive allele of *FAR5*. Specifically, I asked: What are the characteristics of the alpine (*A*) allele compared to the foothill (*F*) allele? Is there any phenotypic effect of the two different alleles of *FAR5*? What are the environmental factors shaping the distribution of the alpine allele of *FAR5* among the populations of *A. arenosa*?

I first demonstrated that natural variation in *FAR5* played a role in adaptation to the alpine environment: I identified four SNPs that were positively selected in the alpine populations across all five high altitude colonizations of *A. arenosa*. Three of these SNPs are coding and are linked together, forming a distinct alpine allele of the *FAR5* protein. Using a unique crossing design based on a natural standing variation of the two identified *FAR5* alleles, I prepared carriers of the non-native alleles on the genomic background of foothill and alpine populations. I then conducted a transplant experiment into the alpine environment of Austrian Alps and examined the phenotypic effect of the two alleles. Due to the reported function of *FAR5* in wax biosynthesis in *A. thaliana*, I used a GC-MS-based metabolite profiling. My results show a clear phenotypic effect of the candidate alpine-adaptive allele. The substrate specificity of the *FAR5* enzyme is shifted from foothill 18C to alpine 16C, resulting in an increased production of a shorter primary fatty alcohol (C16:0-OH) in plants carrying the alpine allele of *FAR5*. This finding experimentally supports *FAR5* as an important player in producing the alpine-adaptive phenotype and brings question about what the function of such metabolomic change is. Thus, I performed a genotype-environment association to unravel which environmental factors might be responsible for the *FAR5* natural variation. I identified an association of the alpine allele of *FAR5* with higher spring precipitation, lower solar radiation in the vegetative season and lower minimum winter temperature. The genome wide association study (GWAS) revealed eight genes significantly associated with similar environmental factors as *FAR5*. Some of these genes are involved in the regulation of seed germination. In connection with the putative effect of the change in substrate specificity on the production of suberin, and *FAR5*'s expression in seed coat, I propose a hypothesis about its adaptive involvement in regulating seed germination. Other factors would also suggest a possible role in pathogen or drought resistance. I suggest that a follow-up targeted study is needed to assess the exact impact of the alpine allele of *FAR5*.

Overall, this study contributes to the understanding of repeated adaptation to the alpine environment through the characterization and functional validation of a candidate alpine-adaptive allele. Unravelling the mechanisms behind such adaptation to dramatic environmental shifts, such as those between lowland and alpine populations, advances our understanding of adaptation. This understanding could be beneficial, for example, in the context of climate change.

**Keywords:** repeated evolution, alpine adaptation, *Arabidopsis arenosa*, *FAR5*, functional genetics, genotype-environment association





# Abstrakt

Opakovaná adaptace představuje jedinečnou příležitost ke studiu mechanismů evoluce v rámci přirozeného experimentu s několika opakováními. Již dříve byla studována opakovaná adaptace *Arabidopsis arenosa*, divokého příbuzného modelového organismu *Arabidopsis thaliana*, na vysokohorské prostředí. Genomická analýza pomocí selekčních skenů odhalila sadu kandidátních alel. Zde uvádím navazující funkční studii kandidátní horské adaptivní alely *FAR5*. Mé konkrétní otázky byly: Jaké jsou vlastnosti horské (*A*) alely ve srovnání s nížinnou (*F*) alelou? Existuje nějaký fenotypový projev těchto dvou různých alel *FAR5*? Jaké faktory prostředí utvářejí distribuci horské alely *FAR5* mezi populacemi *A. arenosa*?

Nejprve ukazuji, že přirozená variabilita genu *FAR5* hrála roli v adaptaci na vysokohorské prostředí: identifikovala jsem čtyři SNP, které byly pozitivně selektovány v horských populacích při všech pěti vysokohorských kolonizacích *A. arenosa*. Tři z těchto SNP jsou kódující a jsou spolu svázány, čímž tvoří jasně odlišitelnou horskou alelu proteinu *FAR5*. Pomocí unikátního křížení založeného na přirozené dostupné variabilitě (standing variation) dvou identifikovaných alel *FAR5* jsem připravila jedince nesoucí nepůvodní alely na genomovém pozadí nížinných a horských populací. Poté jsem provedla transplantační experiment do vysokohorského prostředí rakouských Alp a zkoumala jsem fenotypový projev těchto dvou alel. Vzhledem k uváděné funkci *FAR5* v biosyntéze vosku u *A. thaliana* jsem využila profilování metabolitů pomocí GC-MS. Mé výsledky ukazují jasný fenotypový efekt kandidátní horské adaptivní alely. Substrátová specifita enzymu *FAR5* je posunuta z nížinné 18C na horskou 16C, což vede ke zvýšené produkci kratšího primárního mastného alkoholu (C16:0-OH) v rostlinách nesoucích horskou alelu *FAR5*. Tyto výsledky experimentálně podporují *FAR5* v roli důležitého hráče při produkci horského adaptivního fenotypu a přináší otázku, jaká je funkce takové metabolické změny. Provedla jsem tedy asociaci genotypu s podmínkami prostředí, abych odhalila, které faktory prostředí mohou být zodpovědné za přirozenou variabilitu *FAR5*. Pozorovala jsem asociaci horské alely *FAR5* s vyššími jarními srážkami, nižším slunečním zářením ve vegetačním období a nižší minimální zimní teplotou. Celogenomová asociční studie (GWAS) odhalila osm genů významně spojených s podobnými faktory prostředí jako *FAR5*. Některé z těchto genů se podílejí na regulaci klíčení semen. V souvislosti s možným vlivem změny substrátové specifity na produkci suberinu a s expresí *FAR5* v obalu semen navrhuji hypotézu o jeho adaptivním vlivu v regulaci klíčení semen. Další faktory také naznačují možnou roli v odolnosti vůči patogenům nebo suchu. K posouzení přesného dopadu horské alely *FAR5* je třeba navazující důkladná studie zaměřená na tuto problematiku.

Celkově tato studie přispívá k pochopení opakované adaptace na vysokohorské prostředí prostřednictvím charakterizace a funkční validace kandidátní alely horské adaptace. Odhalení mechanismů zodpovědných za adaptaci na takovéto dramatické změny v podmínkách prostředí, jaké vidíme mezi nížinnými a vysokohorskými populacemi, posouvá naše chápání adaptace. Toto pochopení by mohlo být dále přínosné například v souvislosti se změnou klimatu.

**Klíčová slova:** opakovaná adaptace, horská adaptace, *Arabidopsis arenosa*, *FAR5*, funkční genetika, asociace genotypu a podmínek prostředí (GEA)



# Obsah

Introduction .....	15
Repeated adaptation at multiple levels .....	15
Gene reuse in evolution .....	16
Alpine environment: ‘natural laboratory’ for studying repeated adaptation .....	16
Alpine flora thrives amidst the challenging climatic conditions of the alpine zone .....	16
<i>Arabidopsis arenosa</i> : a model for studying alpine adaptation .....	17
<i>FAR5</i> : know your candidate gene .....	18
Objectives and thesis summary .....	19
Summary .....	19
Methods .....	20
Genomic and transcriptomic datasets .....	20
PicMin and AFD-based selection scan .....	20
AlphaFold .....	21
Cultivation of plants .....	21
Crossing design .....	21
DNA extraction, CAPS .....	22
Transplant experiment, leaf wounding .....	22
Sampling, GC-MS-based metabolite profiling .....	22
Statistical analysis .....	23
Genotype-environment association .....	24
Results .....	25
Four <i>FAR5</i> SNPs show robust and repeated selection signals in alpine adaptation .....	25
Alpine and foothill alleles of <i>FAR5</i> mediate distinct metabolic phenotypes .....	29
Metabolite profiling suggests shift in the <i>FAR5</i> substrate specificity .....	29
The stability of <i>FAR5</i> is not affected by the candidate SNPs .....	32
Environmental drivers of <i>FAR5</i> variation .....	33
Discussion .....	38
<i>FAR5</i> as a candidate gene for alpine adaptation .....	38
The <i>FAR5</i> variation from a phylogenetic perspective .....	38
Repeated positive selection likely acted on standing variation in <i>FAR5</i> .....	39
Multiple non-synonymous SNPs under selection: compensatory evolution, co-evolution or hitchhiking? .....	39
<i>FAR5</i> enzyme in the alpine adaptation .....	40
The role of variation in affecting <i>FAR5</i> properties .....	40

The phenotypic effect of positively selected <i>FAR5</i> mutations .....	41
Role of environmental factors in the <i>FAR5</i> -mediated adaptation to the alpine environment...	42
Environmental drivers of <i>FAR5</i> variation .....	42
Exploring the adaptive role of <i>FAR5</i> .....	43
Conclusions .....	45
Bibliography .....	46
Supplements .....	56





# Introduction

They say variety is a spice of life. From an evolutionary point of view, variation is the very basic aspect of surviving. It enables organisms to react to changes in their environment. Different variants of a phenotypic trait or different genetic variants (alleles further) can be advantageous under different conditions (Blanquart et al., 2013). The more variable a population is, the better it can adapt to environmental changes (Bomblies and Peichel, 2022). The process of filtering alleles in every generation is called natural selection. It is through this process that species become adapted to their environment. Famous example involves three-spined sticklebacks adapting to the freshwater environment via alleles of *Eda* locus available already in the ancestral marine populations (Colosimo et al., 2005).

## Repeated adaptation at multiple levels

In the world around us we can encounter organisms that are strikingly similar without being related. Such organisms are often found in similar environments. That suggests that there are some optimal solutions to problems posed by particular conditions and that the organisms can independently find the optimum. Repeated independent evolution of a particular beneficial trait is called repeated adaptation (Cerca, 2023). Some authors distinguish the modes of repeated adaptation based on relatedness between the organisms as convergent and parallel adaptations (Losos, 2011). But for this distinction we need to know details about the relatedness between the organisms and/or the molecular sources of the repeatedly adaptive mutations. Due to these difficulties, it was suggested to denote all such cases as convergent adaptation in the 2008 (Arendt and Reznick, 2008). But throughout my thesis I will use the term ‘repeated adaptation’ which was more recently proposed as a neutral alternative (Cerca, 2023).

The two main perspectives on the study of the repeated evolution are from the point of view of the phenotype or the genome. Repeated phenotypic evolution refers to the independent emergence of similar traits or characteristics in different lineages of organisms, driven by similar selection pressures. The repeated phenotypic evolution can be exemplified by two plant families *Euphorbiaceae* and *Cactaceae*. Both form succulent forms in the reaction to arid environment and in consequence a layman could easily confuse them because of their phenotypic similarity (Wood et al., 2005). Another famous example is the evolution of a wing in bats and birds in which a same organ was formed through different anatomical changes (Alexander, 2015). In the end the structures serve the same function, to fly.

However, I will focus on examining the repeated evolution from the genomic perspective. Repeated adaptation at the genomic level involves the independent positive selection acting on similar genetic changes in different populations or species, leading to repeated genomic signatures. Such repeated genomic evolution can manifest itself at various levels: the level of molecular pathway, gene or allele. At the molecular level, we see organisms using the same molecular pathways (e.g. adaptation to an extreme abiotic stress in Arctic *Brassicaceae* (Birkeland et al., 2020), same genes (e.g. pelvic reduction via *Pitx1* in vertebrates (Shapiro et al., 2006) or even the same variants (e.g. selection of same SNPs in alpine environment (Bohutínská et al., 2021b) to achieve the repeated adaptation. Here, I focus on repeated use of genes (referred to as ‘gene reuse’ hereafter) and alleles.

## Gene reuse in evolution

Gene reuse is a surprisingly frequent phenomenon (Conte et al., 2012) whose importance in evolutionary biology is recently rising thanks to the development of sequencing and genomic analysis methods. It can emerge essentially in three ways (Lee and Coop, 2019). Firstly, the organisms can repeatedly positively select the variation inherited from their shared ancestors. Secondly, the positively selected variants that we observe as reused are a result of a single mutational event and subsequent spread via gene flow. And finally, the same mutation can originate and subsequently become positively selected independently in multiple lineages (Fraser and Whiting, 2020). Repeatability at the gene level provides insight into the factors influencing the diversity of genetic pathways to adaptation (Yeaman et al., 2018). The fact that a gene has been used repeatedly in the past to solve a particular problem increases the chance that it will be used to solve the same problem in the future.

## Alpine environment: 'natural laboratory' for studying repeated adaptation

The alpine environment represents a natural experiment of adaptation of organisms to a relatively steep gradient of conditions. The character of this experiment encompassing replicates across different latitudes and climatic regions (represented as different mountain ranges) attracted scientists for a long time. The island-like alpine environment provides an opportunity for studying the repeated adaptation to similar selective pressures.

## Alpine flora thrives amidst the challenging climatic conditions of the alpine zone

Defining the conditions of the alpine environment in global terms is challenging because of the variability of various alpine environments. Main factor introducing variability in the definition of the alpine environment is latitude. Indeed, the extratropical and tropical mountain ranges differ in many aspects, one of them being seasonality (Körner, 2022). My study however is limited to the area of Central Europe which simplifies the situation, because across Europe the alpine climate was shown to be similar (Körner et al., 2003).

The climate trends in the temperate zone mountains are long studied and well described. Air temperature is reduced in higher altitudes (Ceppi et al., 2012). The gradient of temperature between foothill and alpine environment is particularly prominent during the growing season (more than during winter) and drives the adaptation of plant traits to cold temperatures (Körner, 2022). Opposite trend is observed in precipitation, where values, but also their fluctuation, increase with increasing altitude (Flohn, 1974). Another important environmental factor is solar radiation. No significant elevational changes of solar radiation were reported in European mountains (Tranquillini, 1960). But under certain sky conditions alpine plants can experience short term extremes, not occurring in foothills. That happens when there is a gap in a thin layer of clouds in the direction of midday radiation, so the plants have to cope not just with the direct radiation but also diffuse radiation from the clouds surrounding the gap. This can be further enhanced by reflectance



from snow (Körner, 2022). The atmospheric pressure decreases with altitude, consequently so does partial pressure of CO<sub>2</sub> (Körner, 2022) which could be of importance for the efficiency of photosynthesis (Singh and Kumar, 1935). The reduced partial pressure of CO<sub>2</sub> can be partially compensated by increased molecular diffusivity (Gale, 1972) however other factors such as the plant's mechanism of CO<sub>2</sub> uptake or air temperature are relevant.

In general, the climate measured by the weather station is not necessarily the climate that alpine plants are experiencing. The specific regional aspects of the environment are extremely relevant from the point of view of a single plant. The topographical properties of the habitat, slope inclination, exposure to sun and wind, and landscape fragmentation by relief can significantly alter the environmental conditions (Körner, 2023). Moreover, the plants themselves can influence the conditions in their favor. By modifying their stature, height above ground, leaf arrangements or the density of the canopy they can buffer the harsh conditions that we experience during our trips to mountains (Körner, 2022). For instance, cushion plants increase the leaf temperature thanks to their compact stature (Salisbury and Spomer, 1964). Overall, the interplay between climate and topography creates in places a very challenging environment for plant survival, nevertheless some plants are able to overcome those challenges.

## *Arabidopsis arenosa*: a model for studying alpine adaptation

The habitats of *Arabidopsis arenosa*, a wild relative of the famous model organism *Arabidopsis thaliana*, span different altitudes which makes it an interesting candidate for studying the alpine adaptation (Knotek et al., 2020; Kolář et al., 2016).

A case of repeated alpine adaptation in *Arabidopsis arenosa* was described by Knotek et al., 2020. They demonstrated that five different lineages of *Arabidopsis arenosa* repeatedly colonized the alpine environment of Alps and Carpathians, representing four to five independent instances of alpine adaptation (Knotek et al., 2020; Wos et al., 2022). After colonization of the alpine environment, *A. arenosa* repeatedly formed a distinct alpine ecotype. To name a few differences between these ecotypes, the alpine plants are shorter and have thicker leaves, bigger and more colourful flowers.

The occurrences of morphologically distinct populations of *A. arenosa* have been reported from four different mountain regions across Europe: Eastern Alps (Melzer, 1960), Eastern (Pachschwöll and Pachschwöll, 2019), Western (Měsíček, J. and Goliášová, K., 2002) and Southern Carpathians (Bartok et al., 2016). In Western Carpathians, both diploid and tetraploid lineages occur (Wos et al., 2019) and are kept separate in the literature about the alpine adaptation of *A. arenosa*. Hereafter, I refer to the five alpine lineages: Niedere Tauern and surrounding foothills of the Eastern Alps (referred to as NT), Rodna Mts. and adjacent regions of Eastern Carpathians (RD), Făgăraș Mts. in Southern Carpathians (FG), Vysoké Tatry Mts. and adjacent foothill diploid populations in Western Carpathians (VT, diploid), Západné Tatry Mts. and adjacent foothill tetraploid populations in Western Carpathians (ZT, tetraploid) as described in (Knotek et al., 2020).

To unravel the processes at the genetic level of adaptation, the genome resequencing and following searches for signatures of positive selection are often taken. Using genome resequencing, a follow up study further found a set of genes showing signatures of repeated selection associated with the alpine environment (Bohutínská et al., 2021b). It found that lineages significantly often reuse these genes and reported a clear pattern: more closely related lineages reused more genes

during alpine adaptation, likely because they share more variation which selection can act on. However, which genes are the most frequently reused among these lineages, what is their role in the alpine environment, and why they repeatedly became targets of natural selection, remained largely unknown.

## *FAR5*: know your candidate gene

In my thesis, I focus on the *FAR5* gene, referred to as fatty acid reductase 5 (TAIR database, Berardini et al., 2015) or fatty acyl-CoA reductase 5 (Domergue et al., 2010). Here I decided to use the name 'fatty acyl-CoA reductase 5' as it better fits the described function of the enzyme.

The family of fatty acyl-CoA reductases (FAR) plays a role in the formation of primary fatty alcohols which are fundamental compounds in suberized tissues (Delude et al., 2016). They catalyze the reduction of fatty acyl-CoA to a primary fatty alcohol during a NADPH-dependent reaction. *FAR5* in particular usually reduces C18:0-CoA to octadecanol (C18:0-OH) via an aldehyde intermediate (Domergue et al., 2010). The FAR enzymes have two main domains: Rossmann-fold domain at the N terminus and FAR\_C domain at the C terminus (Rowland and Domergue, 2012). The Rossmann-fold domain is responsible for binding of NAD(P)H and is common in intermediate short-chain dehydrogenase/reductase proteins (Kavanagh et al., 2008). The FAR\_C domain carries the fatty acyl-CoA reductase activity (Rowland and Domergue, 2012).

It was shown that amino acid changes at the positions 355 and 377 alter the substrate specificity of *FAR5*, resulting in a shift in chain length of substrate and thus the product to 16-carbon chains (Chacón et al., 2013). Therefore, the region between the two functional domains was recognised as being responsible for stability and specificity of the enzyme (Chacón et al., 2013). Throughout my thesis, I use the term 'allele' both to denote a gene variant, but also to refer to the corresponding sequence of amino acids.

In *A. thaliana*, the expression of *FAR5* was reported in roots, wounded leaves and seed coat (Domergue et al., 2010) and also in mature leaves and flowers (Mergner et al., 2020).

As stated above, *FAR5* is involved in suberin biosynthesis. Suberin is a plant biopolymer deposited in the cell walls of certain tissues such as root endodermis, root and tuber peridermis and seed coat (Franke and Schreiber, 2007; Vishwanath et al., 2015). It is a complex heteropolymer composed of polyaliphatic and polyphenolic domain. The polyphenolic domain is anchored in the primary cell wall and mainly consists of polymerized hydroxycinnamates and monolignols (Bernards et al., 1995; Kolattukudy, 2001). The polyaliphatic domain is located between the cell wall and plasma membrane (Woolfson et al., 2022) and is made up from  $\omega$ -hydroxy fatty acids,  $\alpha,\omega$ -dicarboxylic acids, midchain oxygenated fatty acids, unsubstituted fatty acids and primary fatty alcohols (Kolattukudy, 2001; Pollard et al., 2008).

Non-covalently linked root waxes, including besides other compounds also primary fatty alcohols, were reported in association with suberin (Schreiber et al., 2005). They are considered to importantly influence the water diffusion across suberized cell walls (Schreiber, 2010). Suberin and its associated waxes play an important role as a barrier in plant-environment interfaces controlling the movement of water and solutes (Franke and Schreiber, 2007). Suberization was also recognized as essential for wound-healing in plants (Dean and Kolattukudy, 1976). Together with its role in water management it makes suberin, and thus possibly *FAR5*, interesting for the study of environmental adaptation.

# Objectives and thesis summary

In my thesis I first demonstrated that multiple genes show signals of selection in all five alpine lineages of *Arabidopsis arenosa*. I found one of the strongest signals in the gene *FAR5*, which affects the biosynthesis of suberin in *A. thaliana*. I focused on the *FAR5* gene further, asking: Which mutations characterize foothill and alpine alleles of *FAR5*? What is the phenotypic consequence of the alpine allele? Which environmental factors may have triggered the selection acting on the alpine allele? Answering these questions may improve our understanding of the process of repeated adaptation on a gene level.

## Summary

Using the fact that the *FAR5* alleles occur at low frequencies in their non-native environment, I was able to apply a unique crossing design resulting in homozygous rare carriers of the contrasting (alpine and foothill) *FAR5* alleles. Seeds from these crosses were germinated and because they are not artificially modified organisms, I was able to set a transplant experiment in Innsbruck. I planted a set of seedlings in the alpine environment. The set comprised case (alpine plants with foothill allele and foothill plants with alpine allele) and control plants (alpine plants with alpine allele and foothill plants with foothill allele). Based on literature in *A. thaliana* I focused my search for candidate alpine-environment related phenotypes at the level of production of metabolites.

My results from the transplant experiment suggest that there is a phenotypic difference between different alleles of *FAR5* affecting the substrate specificity of the enzyme and the proportions of primary fatty alcohols in the plant tissues. This well corresponds to the proposed function of *FAR5* in production of primary fatty alcohols (Domergue et al., 2010). These results of my functional investigation support the strategy of identifying candidate alleles, which in my case effectively revealed *FAR5* alleles contributing to altered alpine phenotype of *A. arenosa*. However, while my findings underscore *FAR5* not only as a candidate gene for alpine adaptation but also as a determinant of plant phenotype in alpine environments, numerous questions remain unanswered, including the specific functional and fitness effects of alpine alleles.

In the final part of my thesis, I shed light on these questions by demonstrating the correlation between the frequency of *FAR5* alleles within my *A. arenosa* dataset and a set of environmental factors, employing a pRDA analysis. The best explanatory variables suggest influences from early season precipitation, solar radiation in the vegetative season and minimum winter temperature.

The adaptation of organisms to a changing environment presents an ideal opportunity for studying the mechanisms of evolution. Specifically, local adaptation to the alpine environment enables us to study the reaction of an organism to a relatively steep gradient of conditions. By understanding the processes of adaptation of an organism to challenging environments, we may be able to predict certain aspects of adaptive evolution: a knowledge which can be widely applied, for example to weaken the consequences of climate change.

# Methods

## Genomic and transcriptomic datasets

I used a genomic dataset that includes genomes from a total of 73 previously sequenced alpine and foothill populations of *A. arenosa* (Bohutínská et al., 2021b; Konečná et al., 2021; Marburger et al., 2019; Monnahan et al., 2019; Novikova et al., 2016; Preite et al., 2019), mapped to the reference genome of *A. lyrata* (Hu et al., 2011) following the mapping procedure published in (Monnahan et al., 2019). On this data I performed PicMin and AFD-based selection scan (see below).

I annotated each SNP in the genome wide dataset and assigned it to a gene using SnpEff 4.3 (Cingolani et al., 2012) and following *A. lyrata version 2* genome annotation (Rawat et al., 2015). Annotated variants were extracted from vcf format to table using SnpSift, part of SnpEff 4.3, with flags “CHROM POS REF ALT AC AN ‘ANN[\*].HGVS\_P’” and these tables were used as the basis for the subsequent analysis of positive selection.

I also used PacBio HiFi read assemblies from two alpine (lineage VT and FG) and eight foothill individuals of *A. arenosa* (unpublished data kindly provided by Filip Kolář) to assess the relationships between the differentiated candidate SNPs. Specifically, to examine the linkage between the SNPs and check for possible structural variants between alpine and foothill individuals, using BLAST analysis followed by visualization in Geneious. Moreover, these data were used for designing primers and restriction sites (see below).

Furthermore, analysis of *FAR5* gene expression was conducted on data from Wos et al., 2021 where the plants were grown under both the alpine- and foothill-like conditions. The dataset encompassed RNA-Seq data previously generated for a subset of the four alpine and four foothill studied populations (Wos et al., 2021). The edgeR package was used to test for consistent differential gene expression between foothill and alpine individuals, irrespective of their growth conditions (Robinson et al., 2010). Briefly, library sizes (i.e., read counts) were scaled and normalised, dispersion estimated and the ‘glmFit’ function used to test for gene expression differences between alpine and foothill individuals, with treatment as covariate.

## PicMin and AFD-based selection scan

I applied PicMin over 500 bp windows, which is just below the average LD decay of *A. arenosa* (~600 bp, (Bohutínská et al., 2021b)) to test for evidence of repeated genetic differentiation among the foothill/alpine population pairs. PicMin uses order statistics to test whether population genetic summary statistics (in this case  $F_{st}$  (Hudson et al., 1992)) for orthologous genomic regions in different lineages exhibit a common shift towards extreme values in multiple lineages. That indicates the repeated operation of positive selection. PicMin was applied on windows that had data for all five lineages. A genome-wide false discovery rate correction was then performed with a significance threshold of  $q < 0.01$ . In cases of outlier signal spanning adjacent windows, the window with the lowest  $q$ -value and highest  $F_{st}$  was retained.

For the selection scans, the allele frequency difference (AFD), was used as the measure of genetic differentiation (Berner, 2019). The selection scan was performed using the NatGenVarViewer R script ([github.com/mbohutinska/NatGenVarViewer](https://github.com/mbohutinska/NatGenVarViewer)). Briefly, genes were

scanned for outlier single nucleotide polymorphisms (SNPs) differentiated between foothill and alpine individuals. with the outlier AFD cut-off set at  $AFD = 0.6$ . The identification of candidate genes was facilitated by outcrossing in both alpine and foothill populations and following high nucleotide diversity that aids candidate gene detection in *A. arenosa* (Yant and Bomblies, 2017).

## AlphaFold

To visualize the changes between the A and F variant of FAR5 I used AlphaFold v2.0 (Jumper et al., 2021). I used the protein sequence of *A. lyrata*. For highlighting the relevant structures, I gave different colors to different functional domains as described in (Chacón et al., 2013). Blue was used for the Rossmann-fold domain, green for the FAR\_C domain and gray for the part of protein with SNPs affecting the specificity and stability of the enzyme. Furthermore, I also highlighted the molecular surface of the residues in *A. lyrata* at the positions corresponding to the candidate SNPs.

## Cultivation of plants

For the experimental part, I used two different alpine lineages of *A. arenosa*, one diploid VT and one tetraploid ZT, as replicates. I selected 4 populations (2 pairs of alpine and foothill) from our seed collection, based on their genotype frequencies of the three differentiated coding FAR5 SNPs (selecting the purest alpine and foothill populations, which segregate FAR5 SNPs). For alpine populations I used seeds from TKO (ZT lineage, 1783 masl) and ZEP (VT lineage, 1625 masl) and for foothill populations HRA and SUB (further details regarding these populations see Suppl. Table 1). I germinated the seeds in the walk-in chamber provided by PSI under controlled conditions (23/18 °C, 16/8 h day/night, light intensity 150  $\mu$ E). I vernalized the seedlings for 4 weeks (6 °C, 8/16 h day/night, light intensity 150  $\mu$ E). After vernalization I transferred them to conditions (20/15 °C, 16/8 h day/night, light intensity 150  $\mu$ E) for maturing and subsequent crossing.

## Crossing design

To generate individuals with contrasting combinations of genomic background and *FAR5* allele, I genotyped the plants for the presence of alpine and foothill alleles, using restriction fragment length polymorphism in the *FAR5* locus (method CAPS, see below). I first aimed to identify rare heterozygous carriers of the contrasting allele. I then reciprocally crossed these heterozygotes within the population (aiming at three pairs of reciprocal crosses within each population), resulting in a F1 generation equally segregating the alpine and foothill allele (1AA:2AF:1AA in diploids, more complex pattern of genotypes in tetraploids, with most common being 1AAAA:8AAAF:18AAFF:8AFFF:1FFFF) (Fig. 5A). I did another round of crossing to generate *FAR5*-homozygous (or, in case of tetraploids, nearly homozygous) F2 generation which were transplanted to an alpine environment (see below).

I managed to successfully apply this design on three of the four populations originally selected for the experiment. The alpine population from the VT lineage did not show sufficient frequency of the foothill allele, therefore I only used the results from the two populations from the ZT lineage (foothill HRA and alpine TKO) and the foothill population (SUB) from the VT lineage in the subsequent analysis.

# DNA extraction, CAPS

I extracted DNA as described in Supplementary Method 1. DNA samples were stored at 5 °C in the TE buffer for further use. Before PCR I diluted the sample to 10 ng/μl.

For assessing the allele in an individual, I used CAPS (cleaved amplified polymorphic sequences), a method where PCR-amplified DNA fragments are digested by a restriction enzyme. The restriction fragment length polymorphisms are then visualized on an electrophoretic gel (Konieczny and Ausubel, 1993). Specifically, the primers for amplification were designed in Geneious Prime 2021.2.2 software. The PCR mixture contained 10 ng of plant genomic DNA, 5 pmol of the F primer (TCATGTTGACAGATACCAGTGGA), 5 pmol of the R primer (CCTGGAGAGGTAGTTTGTAACGT), 2 μl of MyTaq reaction buffer (containing dNTPs, MgCl<sub>2</sub> and enhancers) and 0.5 U of Taq polymerase in a total volume of 10 μl. PCR was performed by Eppendorf Vapo.protect Mastercycler Pro for 1 min at 95 °C, 30 cycles of 15 s at 95 °C, 15 s at 60 °C and 10 s at 72 °C, followed by 5 min at 72 °C and finally had been held at 10 °C. The resulting fragment had 283 bp.

For digestion of the alpine allele, I used the DdeI restriction enzyme (Howard et al., 1986). This enzyme was designed in Geneious Prime 2021.2.2 software using alpine and foothill sequence of *FAR5* assembled from our long-read resequencing dataset. The enzyme was designed to cut at the position 377 that is in the middle position among my differentiated SNPs and also is the most foothill-alpine differentiated SNP. The reaction mixture for enzyme digestion contained 2 μl of the PCR product, 6.7 μl of ddH<sub>2</sub>O, 0.3 μl of the restriction enzyme (1000 U/ml) and 1 μl of 10× enzyme buffer, which was incubated at 37 °C for 15 min, 65 °C for 30 min and held at 10 °C in the end. The products were visualized by 1.8% agarose gel electrophoresis run at 120 V for 30 min. The *F* allele had length of 283 bp and the *A* allele was cleaved into two fragments of 69 bp and 214 bp. The agarose gel electrophoresis results were photographed using the GelDoc Go imaging system. I manually scored each genotype, using band intensities to distinguish the different classes of heterozygotes (allele dosage) in tetraploids (Gebhardt, 2007).

# Transplant experiment, leaf wounding

Four weeks after germination, F<sub>2</sub> seedlings were transplanted into pots and burrowed in sand in the Alpengarten Patscherkofel (1960 m a.s.l.). After two months I wounded the leaves by applying pressure by the tweezers for a few seconds (Sözen et al., 2020). I then left the lesion to heal and harvested the leaves after 5 days. At this time point there is expected to be the highest expression of *FAR5* in *A. thaliana* (Domergue et al., 2010) and also my test runs determined it as the best time to observe the effect of the wounding (Suppl. Fig. 1).

# Sampling, GC-MS-based metabolite profiling

For root sampling, plants were removed from the pots, roots were carefully cleaned in water and briefly dried them with paper towels. Afterwards, a 5 cm root segment was collected 1 cm below the incipience of the leaf rosette and placed in liquid nitrogen (-196 °C). Both control and wounded leaves were sampled after 5 days of the wounding treatment, cut and immediately frozen in liquid

nitrogen. All the samples were freeze-dried with Zirbus VaCo 2 for three days and then stored at -80 °C.

The preparation of samples for GC-MS followed a slightly modified version of the non-extraction method for global-acyl-chain profiling previously described in (Delude et al., 2017), during which the samples are directly depolymerized by acidic transmethylation. For the method, ca. 3 mg DW of root material and 5 to 6 mg DW of leaf material were used. The aliquots of freeze-dried material and quality controls were extracted at 85 °C for 3h in 1 mL of 5% (v/v) sulfuric acid in methanol containing heptadecanoic acid (C17:0), pentadecanol (C15:0- OH) and C15-hydroxypentadecanoic acid ( $\omega$ -OH-C15:0) as internal standards. After adding 1 mL of 2.5 % (w/v) NaCl and 2.2 mL of methyl tert-butyl ether (MTBE) and centrifugation I collected the upper phase. 1mL of 100 mM Tris base pH 8.0 containing 0.09% (w/v) NaCl were added to the MTBE phase before proceeding with another centrifugation. The upper phase was collected again, and MTBE evaporated under a gentle stream of nitrogen prior to derivatizing the samples using 100  $\mu$ L of 99% BSTFA (N,O-Bis(trimethylsilyl)-trifluoroacetamide) with 1% TMCS at 110 °C. Finally, after evaporating the solvent, I dissolved the products in 500  $\mu$ L of heptane:toluene (1:1, v/v).

Metabolites were separated on a 30 m BPX70 column with 0.25 mm ID and 0.25  $\mu$ m film thickness from SGE (Melbourne, Australia) using a Trace 1300 gas chromatograph coupled to a TSQ 8000 triple quadrupole mass spectrometer from Thermo Scientific (Massachusetts, USA) with a Topaz 4.0 mm ID Single Taper Inlet Liner w/Wool from RESTEK (Bad Homburg, Germany). A 1  $\mu$ l aliquot of the sample was injected in splitless mode for GC-MS analysis. The temperature of the injector was held at 250 °C. The column oven temperature was held at 50 °C for 1 min and then increased from 50 °C to 200 °C at a rate of 25 °C per min, followed by a 1 min hold, and then was ramped up again at a rate of 10 °C per min to a final temperature of 320 °C, which was held for 8 min. The total run time was 28 min. High purity helium was used as the carrier gas at a flow of 1.5 ml per min. A mix of alkanes with 10 to 36 carbons was injected in the middle of the sequence for retention indices calibration.

Compound spectra were extracted from the raw data files using the “Automated Mass-spectral Deconvolution and Identification System” (AMDIS) and compared against a custom-built mass spectral library and the commercial library of the National Institute for Standards and Technologies (NIST) using the MS Search software (v2.4). Compounds identification were based on both spectral and retention index match. Afterwards, I used Xcalibur software (v4.5, Thermo Scientific) to determine peak areas for compound-specific fragments for relative quantification of identified compounds in the biological samples.

The final data I used for wounded leaves is in Suppl. Table 5 and for roots in Suppl. Table 6.

## Statistical analysis

For metabolite profiling data I used Wilcoxon rank sum test to compute pairwise comparisons in the relative abundance of the compounds between different genotypes. I used Kruskal-Wallis rank sum test to test if there are significant overall differences between the groups. This statistical analysis was performed using the stats package in R version 4.1.2 (R Core Team, 2021). All main data handling was conducted with this software. In the analysis of environmental data, I used the function `ggpairs` from R package `GGally` (Schloerke et al., 2021) to compute and plot Pearson correlation coefficient. For creating the illustrational maps I used features from R packages: `giscoR`,

elevator, sf, terra and raster (Hernangómez, 2024; Hijmans, 2023a, 2023b; Hollister et al., 2023; Pebesma, 2018).

## Genotype-environment association

I used redundancy analysis (RDA, implemented in R package *vegan* (Oksanen et al., 2022)) to identify the relationship between the frequency of *A* allele of *FAR5* and environmental variation. RDA is a powerful tool based on multivariate regression which enables us to model linear relationships between genomic data and environmental variables. More precisely, I used a derived method called partial redundancy analysis (pRDA) that allows us to look separately at the explained variability for different explanatory variables and to account for the correlations between the variables. I followed a similar course of actions as described in (Capblancq and Forester, 2021). For this analysis I standardized all the variables, i.e. subtracted the mean and divided by the standard deviation. To account for the correlation between the environmental variables, I used the `ordi2step` function from the *vegan* package in R. The method starts with an empty model and subsequently compares the current model with the global model (all variables) after adding other environmental variables one by one. Once the model cannot get better than the global model, the p-value reaches the threshold or the adjusted  $R^2$  decreases, the method stops. After each round the method removes the effect of the best explanatory variable and thus all the variables that strongly correlate with it. As a consequence, one can find the second best explanatory variable without the effect of the first best. Genomic data were from the dataset mentioned above. Values of 103 environmental variables for all populations were extracted from the WorldClim database with a resolution of 2.5 min (Fick and Hijmans, 2017). For the final analysis I used monthly data for minimum temperature (`tmin`), maximum temperature (`tmax`), wind speed (`wind`), solar radiation (`srad`), precipitation (`prec`) and water vapor pressure (`vapr`). I also used specific value for each population corresponding to the most important vegetative month based on our field data for six variables mentioned above and average temperature (`tavg`). My full RDA model comprised 79 environmental variables.

For conducting GWAS I followed (Capblancq and Forester, 2021). I assessed to which extent allele frequency at each SNP corresponds to the set of environmental conditions shaping the *FAR5* allele frequencies across all populations of *A. arenosa*. Specifically, I used the following model: `RDA_env <- rda(AllFreq ~ prec_May + srad_vegMonth + tmin_Feb + Condition(PC1 + PC2), Variables)`. Where `AllFreq` are population allele frequencies calculated across all non-synonymous SNPs found in all 73 populations in my genomic dataset (see above). For assessing the neutral genetic structure (PC1 and PC2) I used 10 000 SNPs randomly chosen from all the synonymous SNPs. I selected all SNPs with p-value lower than  $1 \times 10^{-50}$  for further interpretation.



# Results

## Four *FAR5* SNPs show robust and repeated selection signals in alpine adaptation

To identify genes showing signatures of selection associated with alpine environment, I used a collection of previously published genomes of *A. arenosa* and selected 299 individuals of all five lineages where colonization of the alpine environment has been described (73 VT, 75 ZT, 89 NT, 40 FG and 22 RD, Fig. 1A). This involved 87 alpine and 212 foothill individuals.

To detect genes showing signatures of repeated positive selection I applied PicMin, a method that can test for repeated genetic differentiation, using the results of genome scans (Booker et al., 2023). To do so, I compared genetic differentiation between each of the five alpine populations and their foothill counterparts from the same mountain range (Fig. 1A) by calculating  $F_{st}$  in 500 bp genomic regions. Using PicMin, I next identified 285 significantly differentiated windows ( $q$ -value  $< 0.05$ ) and 100 of these overlapping with 30 genes were found to be differentiated in all 5 lineages, likely contributing to an aspect of presumably highly polygenic alpine adaptation (Suppl. Table 2).

At the minimum  $q$ -value of 0.0097 (given the number of windows and replicates), I identified 79 windows showing signatures of repeated selection and 41 of these windows in all five lineages. These 41 windows overlapped with 13 genes, 6 of which have well described protein functions in *A. thaliana*. Specifically, these are photoreceptor phytochrome B involved in germination, myrosinase TGG1 producing compounds toxic to herbivores, RPM1 conferring a resistance to certain pathogens, a transporter ACD11, LAZ5 responding to pathogens and fatty acyl-CoA reductase *FAR5* involved in formation of suberin (TAIR database, (Domergue et al., 2010)). The *FAR5* gene showed a particularly strong signal of repeated differentiation at SNP level (Fig. 1D) and therefore I studied it further.

Calculation of allele frequency difference (AFD) between a set of all alpine and all foothill populations in my dataset revealed that four SNPs drove the strong differentiation pattern in the *FAR5* window (Fig. 1D). These involve three amino acid changes, all located outside of the two conserved *FAR5* domains (Fig. 1CE), and one intron variant. The most alpine-differentiated SNP is located at position 377 (Val377Leu, AFD = 0.68), a site demonstrated to alter *FAR5* substrate specificity in *A. thaliana* (Chacón et al., 2013), which was therefore my primary center of interest for the downstream analyses. The other two amino acid changes involved Ala384Ser, located in a region possibly affecting *FAR5* stability, and Cys329Gly (Fig. 1CE).

To better understand the genetic variability in the *FAR5* locus, I next used PacBio HiFi read assemblies from two alpine (lineage VT and FG) and eight foothill individuals of *A. arenosa*. All three alpine-characteristic amino acids were linked together, making either fully alpine allele (329Gly - 377Leu - 384Ser; *A* allele hereafter), or fully foothill allele (329Cys - 377Val - 384Ala; *F* allele hereafter, Fig. 1D). The coding SNPs were not linked to any structural variant in the *FAR5* coding region, suggesting that these SNPs alone primarily underlie foothill-alpine protein differences (Suppl. Fig. 2).

The comparison of the 10 *A. arenosa* *FAR5* assemblies to the *A. thaliana* and *A. lyrata* *FAR5* reference sequences revealed that for the position 329 the change to glycine is likely a reversal into

an ancestral state. The ancestry at the position 377 is more ambivalent; however, the change to leucine is in concordance with the *A. lyrata* state (Fig. 1E). The apparent evolutionary novelty of the third coding SNP, together with the fact that the substituted amino acids (Ala384Ser) have different chemical properties, and position in the region known to be involved in FAR5 stability, makes it an additional candidate SNP of interest. I further found that while coding *FAR5* regions are conserved (average pairwise alignment identity, PAI = 98.5 %), introns are considerably more variable (PAI = 91.4 %). In accordance with this observation, I found that three small deletions are linked to the alpine-characteristic SNP in the intron 8 (Suppl. Fig. 3).

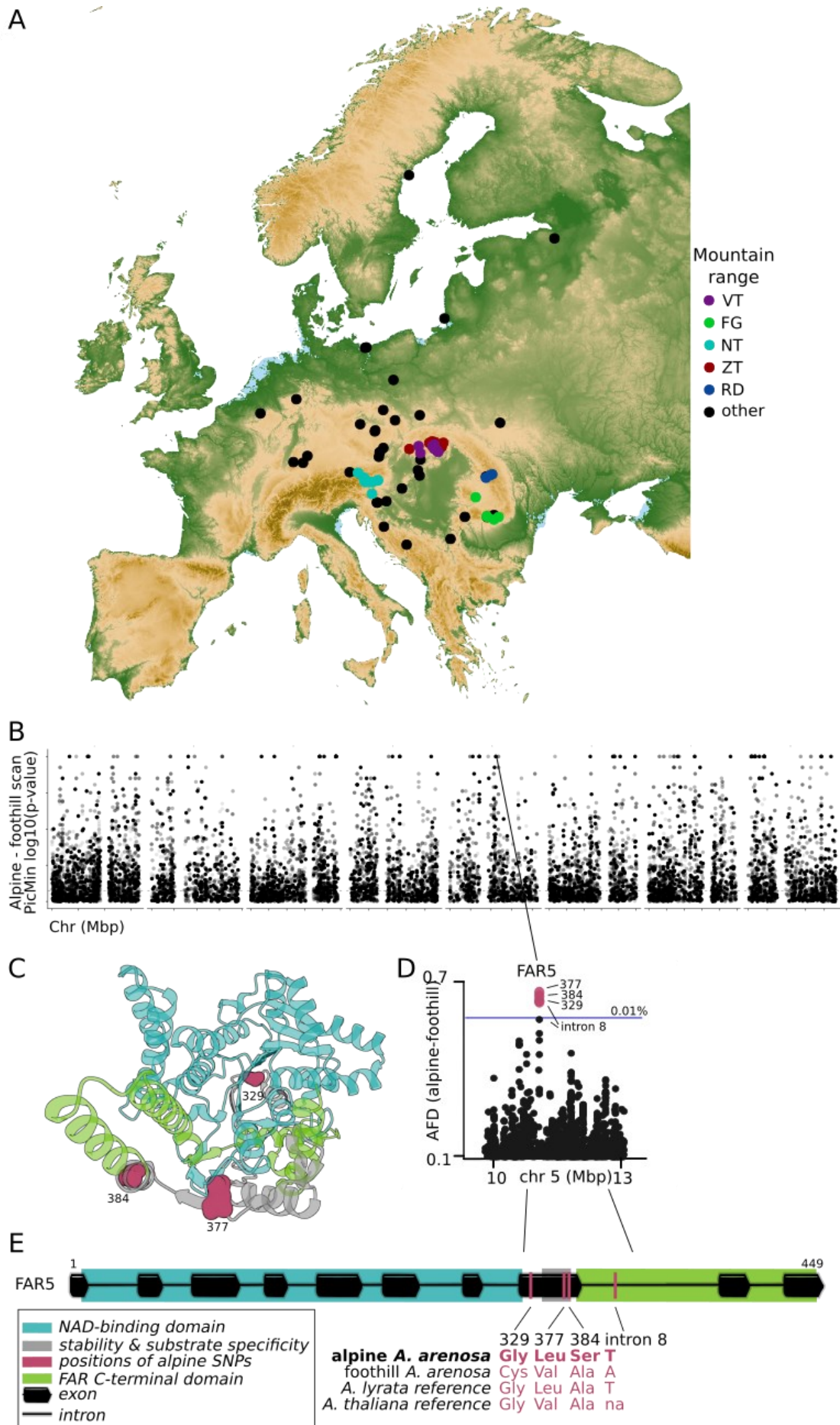


Fig. 1: Signals of positive selection acting on FAR5 in *A. arenosa*. A: the genomic dataset of 73 populations of *A. arenosa*, alpine and foothill populations from the five alpine lineages are marked by colored dots, other populations are black; B: Manhattan plot

depicting the results of *Picmin* across eight chromosomes of *A. arenosa*, *FAR5* overlaps with one of the peaks showing the strongest signal of repeat positive selection; C: AlphaFold-visualized protein structure of *FAR5*. Highlighted are functional domains and the molecular surface of the three coding candidate SNPs; D: zoom-in into the *FAR5* region on the chromosome 5 showing high allele frequency difference (AFD) for three coding and one intron SNP; E: gene model of the *FAR5* gene and phylogenetic comparison of SNPs with the closest relatives. Candidate SNPs highlighted by vertical red lines, functional domains by colorful boxes.

Using AlphaFold protein structure predicted for *FAR5* from *A. lyrata*, I found that the molecular surface of the (alpine) leucine residue at position 377 is predicted to extend towards the NAD-binding domain of *FAR5* (Fig. 1C). This may suggest that the alpine *FAR5* allele evolved to alter substrate specificity of *FAR5*, and in consequence the composition of fatty alcohols in alpine plants.

To assess if the SNPs and deletions in the intron 8 could have an impact on *FAR5* protein through its retention, I mapped 96 available leaf transcriptomes of *A. arenosa* on a foothill *FAR5* gene assembly from the VT lineage. I have not detected any sign of intron 8 retention, but I noticed slight differences in transcription around intron/exon boundaries, which could make *A. arenosa*'s *FAR5* protein few amino acids different compared to the one in *A. lyrata*. Finally, I found slightly higher levels of *FAR5* leaf transcription in alpine compared to foothill individuals, which could make an effect of *FAR5* allele in alpine populations more pronounced (Fig. 2).

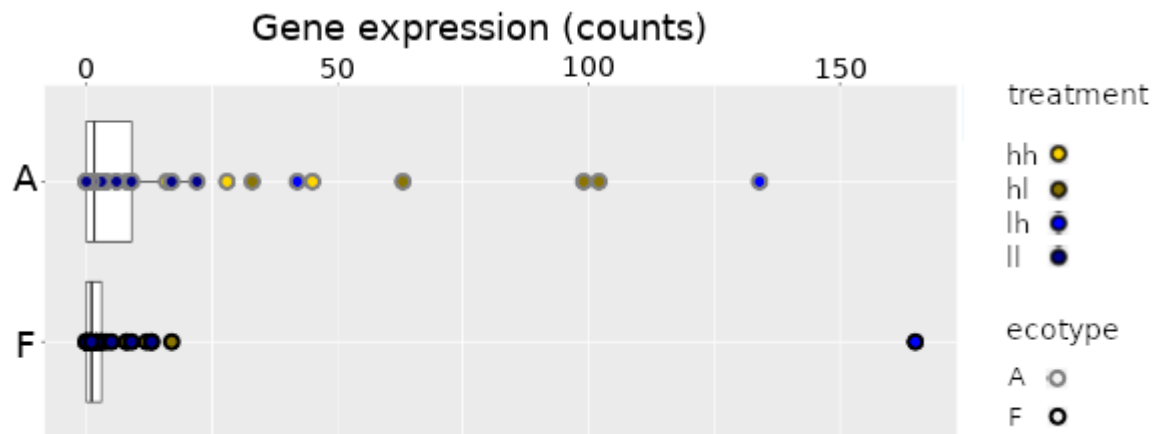


Fig. 2: The degree of leaf *FAR5* expression in alpine and foothill individuals. Blue and yellow highlight different treatments under which the plants were grown.

To understand the evolution of the three candidate coding SNPs from a broader phylogenetic perspective, I explored *FAR5* sequences across 25 million years of evolution of *Brassicaceae*. To do so, I gathered *FAR5* protein sequences from 30 unique species from the *Brassicaceae* family, aligned them and explored how much my candidate positions vary across evolution, how unique are the alpine amino acids and how unique is the biochemical impact of the foothill-alpine change.

The position 329 was not highly conserved throughout the *Brassicaceae* family (PAI = 54.3 %) with glycine as the most abundant amino acid at this position (Fig. 3). I do not observe any change to cysteine in the 30 sequences which makes the protein corresponding to the *F* allele exceptional. Nevertheless, I noticed changes to other hydrophilic amino acids - serine and asparagine. Further, the position 377 does not show a pattern of strong conservation either (PAI = 55.2 %). The foothill version of 377Val is prevalent in the family. The *A* allele variant 377Leu corresponds to *A. lyrata* and to three other species - *Capsella rubella*, *Capsella grandiflora* and *Camelina sativa*. Lastly, the serine at position 384 is different from all the other sequences in my *Arabidopsis* dataset. In general, this position seems rather conserved throughout *Brassicaceae* (pairwise alignment identity at position 384 = 80.7 %), with alanine being the most prevalent amino acid. I observed three instances of amino acid change at this position, one of them being a change to serine in *Eutrema salsugineum*. The other two changes were to threonine in *Lepidium sativum* and to glycine in

*Camelina sativa*. Overall, biochemically most relevant are the amino acid changes at positions 329 and 384 which also show a certain uniqueness across the *Brassicaceae* family.

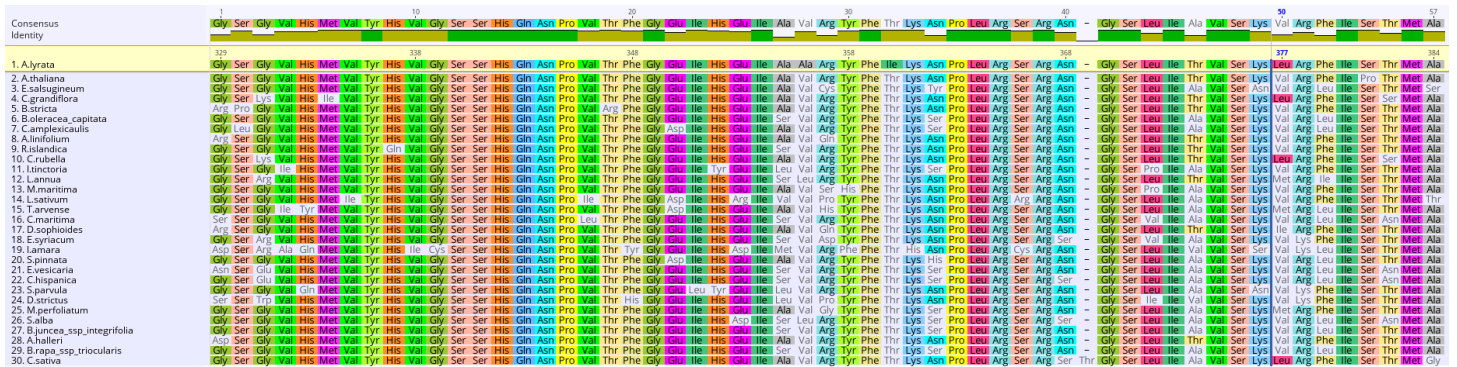


Fig. 3: Protein sequence alignment of 30 *Brassicaceae* sequences of *FAR5*, shown are protein positions 329-384 according to *A. lyrata*, encompassing all three candidate coding SNPs (positions 329, 377 and 384).

## Alpine and foothill alleles of *FAR5* mediate distinct metabolic phenotypes

The alpine adaptation is deeply intertwined with the environmental conditions. Because of the possibility that *FAR5* alleles are both transcriptionally and functionally manifested in an environment-dependent manner, I decided to determine the effect of carrying *A* or *F* *FAR5* allele in natural alpine conditions. Hence, I conducted a transplant experiment into an alpine environment, using alpine and foothill individuals, crossed to be homozygous for either *A* or *F* *FAR5* allele. To prepare such plants, I implemented a crossing design utilizing the natural variation of *A. arenosa* (Fig. 5A). In 78 % of foothill and alpine populations, rare non-native alleles occur (AF of *A* allele across all foothill populations = 0.24, AF of *F* allele across all A populations = 0.15, Fig. 10A). I thus performed a large-scale genotyping of alpine and foothill populations from VT and ZT regions and identified heterozygous plants carrying these rare contrasting variants. I crossed those within populations, genotyped the resulting F1 families and crossed the individuals carrying the non-native allele, resulting in a set of rare allele-homozygous lines (foothill plants homozygous for the *A* allele and likewise for plants originally from the alpine population) and control plants (foothill plants homozygous for the *F* allele and vice versa; Fig. 5A). Such natural variability-based crossing design allowed me to test the effect of both alleles on their native genome background (as compared to for example transforming the *F* and *A* alleles into model species *A. thaliana*) and also enabled me to transplant them to the actual alpine environment, which would not be possible with genetically modified plants.

## Metabolite profiling suggests shift in the *FAR5* substrate specificity

The positions of candidate SNPs within *FAR5* stability and substrate specificity-related domain offer two hypotheses regarding their specific functional impacts: (1) they affect *FAR5* substrate specificity and in consequence the relative abundance of the products or (2) they affect *FAR5* stability and in consequence the accumulation of *FAR5* products. Together with the proposed



function of FAR5 enzyme in suberin biosynthesis, this led me to focus my inquiry about the phenotypic effect of the allele on the production of metabolites, especially on 16- and 18-carbon chain metabolites. I presumed the *F* allele to produce the octadecanol (C18:0-OH) and the *A* allele to produce hexadecanol (C16:0-OH) based on their sequence similarity to *A. thaliana* variants from Chacón et al., 2013 (Fig. 5D). Based on the expression patterns of FAR5 in *A. thaliana* (Domergue et al., 2010; Klepikova et al., 2016; Mergner et al., 2020), I primarily focused on the composition of wax layers of roots and wounded leaves.

To see if any differences in the GC-MS (gas chromatography-mass spectrometry, see Methods) quantification of these metabolites were not given just by different overall metabolic activity of the two alleles, or if there are not genetic variants linked to FAR5 allele which would affect the overall spectrum of metabolites, and in consequence biasing my analysis of FAR5-dependent metabolite changes, I started by analyzing a set of neutral metabolites (metabolites except for hexadecanol, octadecanol and their related compounds). I analyzed this set of neutral metabolites across my 46 experimental individuals and asked if there are any signals of clustering by the allele. PCA of all neutral metabolites from wounded leaves does not show any significant effect of the *FAR5* alleles (Fig. 5B). The metabolites do not cluster neither by the allele nor by their original ecotype. Bearing this in mind, I continued by analysis of the relevant compounds according to my hypothesis about *FAR5* function.

Comparison of individual compounds between individuals carrying *A* and *F* allele has shown that the individuals with *A* allele produced proportionally more hexadecanol (C16:0-OH) than octadecanol (C18:0-OH) compared to the individuals with *F* allele, independent of their alpine and foothill origin ( $p < 0.001$ ,  $N = 61$ , Kruskal-Wallis rank sum test; Fig. 4 left). This pattern is consistent across all three populations in my experiment and in both my measured tissues. In roots the pattern is more prominent and the proportional production of hexadecanol is significantly higher in all compared populations ( $p = 0.004/0.01/2.027e-05$ ,  $N = 11/22/28$  for VT, ZTA and ZTF, respectively, Fig. 5C).

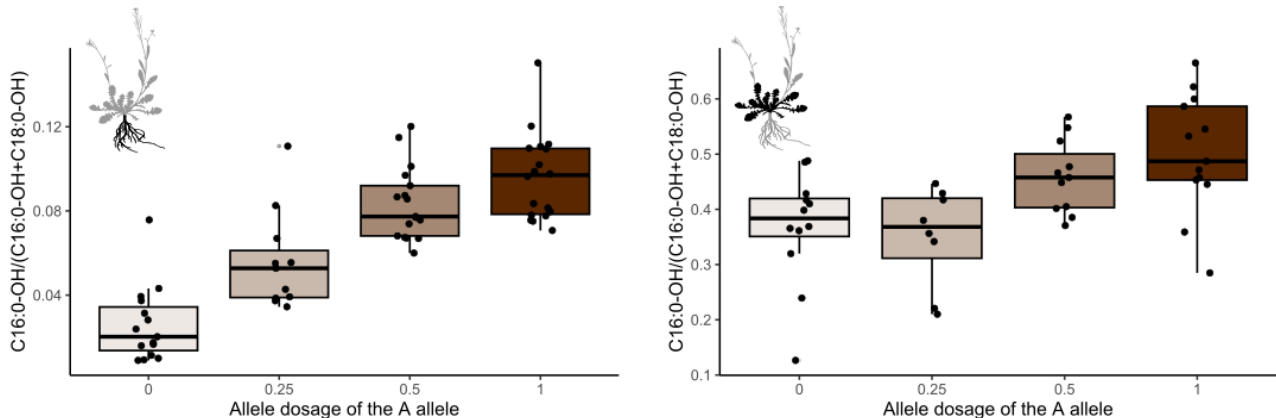


Fig. 4: Proportion of hexadecanol (C16:0-OH/C16:0-OH+C18:0-OH) in roots (left) and wounded leaves (right) in relation to the dosage of the *A* allele of *FAR5* across all samples. The browner the box is, the more *A* alleles the individuals possessed. 0 – the individual had a genotype of *FF* or *FFFF*, 0.25 – *AFFF* genotype, 0.5 – *AF* or *A AFF* genotype, 1 – *AA* or *AAAA* genotype.

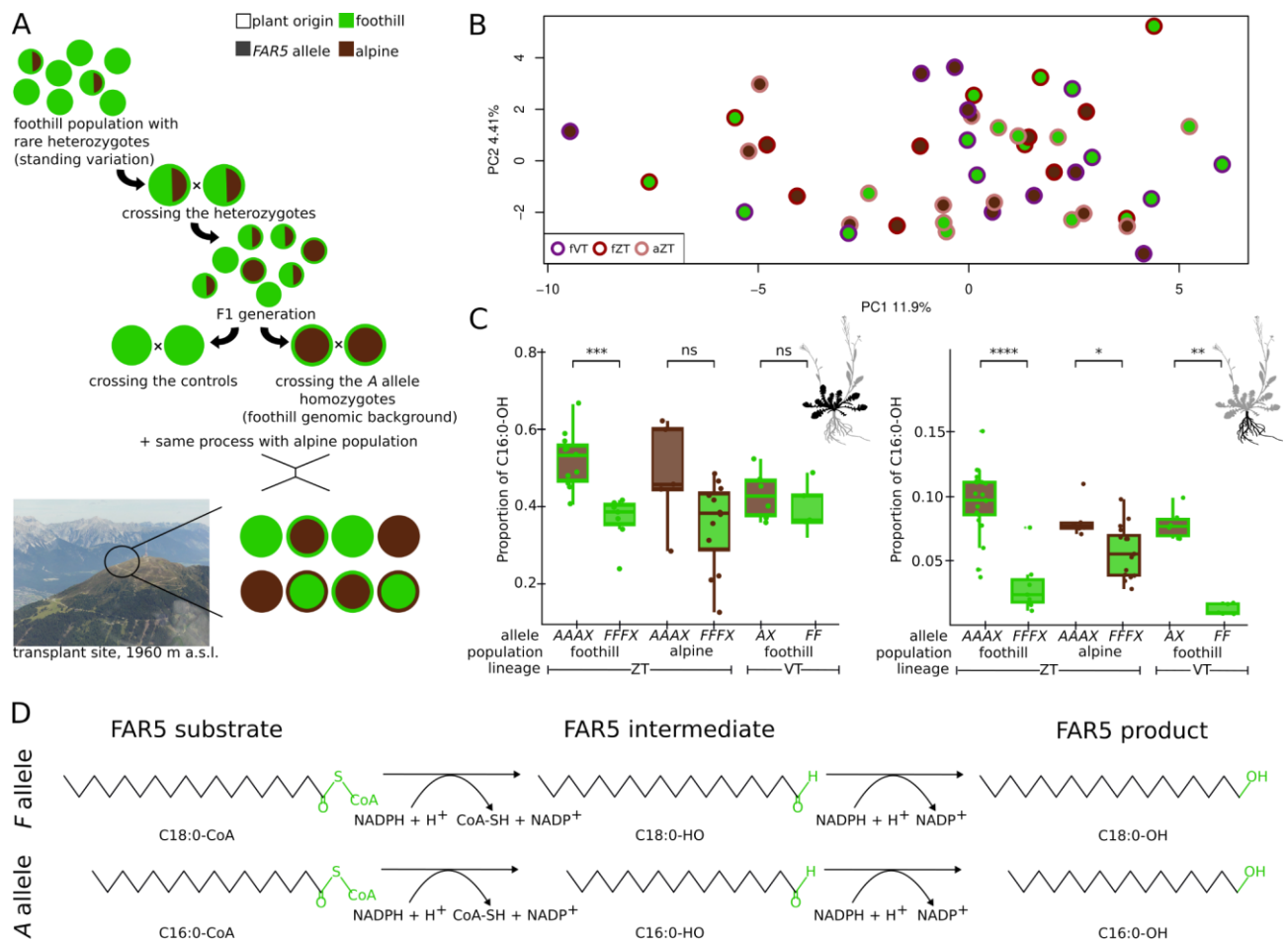


Fig. 5: Metabolomic characterisation of the FAR5 alleles in the natural alpine environment; A: design of crossing the FAR5 genotypes for transplant experiment and the resulting combinations of genotypes; B: PCA of 'neutral' metabolites (see text for details) from the wounded leaves across all individuals in my metabolomic analysis; C: proportion of hexadecanol (C16:0-OH/C16:0-OH+C18:0-OH) in wounded leaves and roots, X allele can be either A or F due to imperfect crossing; D: my hypothetical model of phenotypic effect of FAR5 A allele. Color legend for all figures: green = foothill, brown = alpine, outline = original ecotype, fill = FAR5 allele.

For wounded leaves I firstly tested if leaf wounding had any effect on the production of C16:0-OH and C18:0-OH in my populations of *A. arenosa*. Indeed, the inquiry of the metabolites revealed a significant difference in the relative abundance of hexa- and octadecanol between the unwounded control leaf and a leaf harvested five days after wounding across all the samples ( $p < 0.001$  in both cases,  $N = 46$ , Wilcoxon rank sum test, Fig. 6).

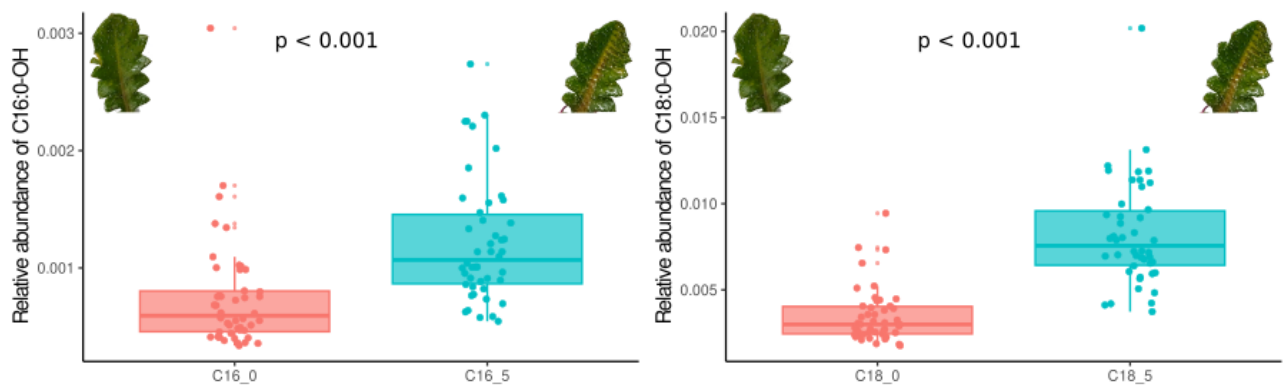


Fig. 6: Relative abundance of hexadecanol (left) and octadecanol (right) in the unwounded control leaves (red) and in the leaves 5 days after wounding (blue).

Specifically, I observed the increase in the production of both fatty alcohols in the wounded leaves. I thus asked if the degree of induction is allele-dependent. I expected the plants with *A* allele to have a higher increase in the production of hexadecanol than plants with the *F* allele and plants with the *F* allele were expected to increase production of octadecanol more than plants with the *A* allele. Indeed, the comparison of the differences in relative abundances of hexadecanol and octadecanol show that the induction of production of C16:0-OH by wounding is higher in the plants with *A* allele, and induction of production of C18:0-OH by wounding is higher in the plants with the *F* allele (Fig. 7).

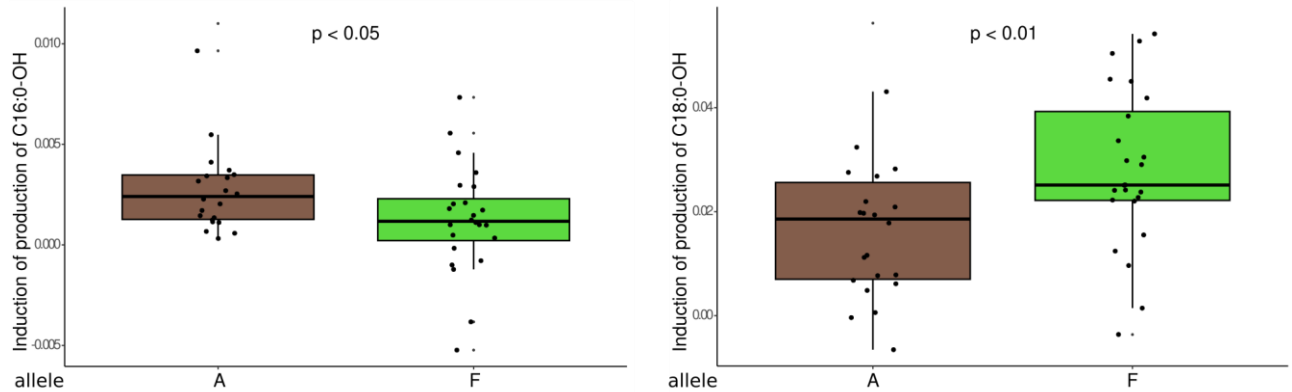


Fig. 7: Induction of hexadecanol (left) and octadecanol (right) production (difference between unwounded and wounded leaf amounts) based on allele; green=foothill, brown=alpine

Because the unwounded control leaves varied in the relative abundance of both fatty alcohols, for the final analysis, I standardized the wounded samples by their unwounded controls. This allowed me to test the effect of the allele on the fraction of metabolites induced by wounding, which is directly linked to *FAR5* expression. When comparing the proportion of hexadecanol across all my populations, I found that plants carrying the *A* allele produced more hexadecanol ( $p=0.002$ ,  $N=44$ , Kruskal-Wallis rank sum test, Fig. 4 right), independent of their foothill or alpine origin. That is consistent with the pattern observed in roots. In wounded leaves, I found significant increase in the C16:0-OH proportion in the foothill population from ZT lineage ( $p\text{-value} = 0.0003$ ,  $N=18$ ) and similar but nonsignificant increase in the other two lineages ( $p = 0.4/0.1$ ,  $N = 11/17$  for VT and ZTA, Fig. 5C).

Other evaluated fatty alcohols with different chain lengths did not show consistent and strongly significant differences between the alleles neither in roots nor in the wounded leaves (Suppl. Fig. 4-9). All in all, these results suggest that the changes between the *F* allele and the *A* allele alter the substrate specificity from 18C to 16C of the *FAR5* protein.

## The stability of *FAR5* is not affected by the candidate SNPs

Two main characteristics of an enzyme are its specificity and its stability. After finding out that the specificity of *FAR5* is shifted if the *A* allele is present, I next tested if there is any allele-dependent effect on *FAR5* stability. As enzyme stability correlates with its efficiency, I compared the sum of the two products of *FAR5*, octadecanol and hexadecanol (corresponding to the overall efficiency of *FAR5*, independent of its specificity (Kunka et al., 2023)). There were not any consistent differences between individuals carrying the *A* and *F* allele in the sum of products. There were two significant, yet opposite, differences in the sum of the two products between my test groups: an increase in individuals with *A* allele in ZTF population ( $p=0.027$ ) and decrease in the individuals with the *A* allele in ZTA population ( $p=0.0024$ ) both in root samples. The rest of the comparisons



were not significant ( $p > 0.05$ ). Therefore, I conclude that amino acid changes in the *A* allele of *FAR5* affect the substrate specificity, while the stability of the protein is not affected (Fig. 8).

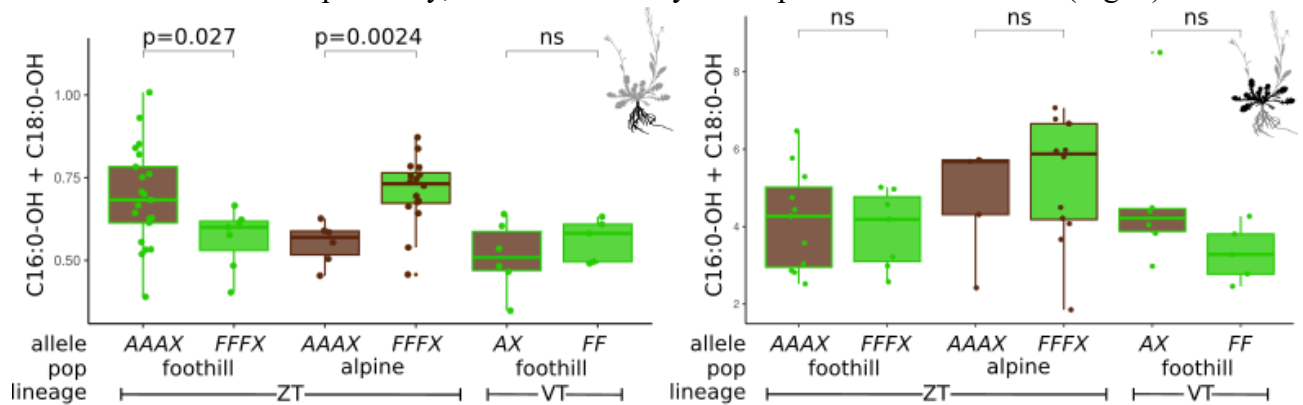


Fig. 8: The sum of relative abundance of hexadecanol and octadecanol, which I use as a proxy of *FAR5* stability, for the samples from roots (left) and wounded leaves (right); ns: non-significant result ( $p$ -value  $> 0.05$ ), *X* allele can be either *A* or *F* due to imperfect crossing; box outline=ecotype, fill=allele, green=foothill, brown=alpine

## Environmental drivers of *FAR5* variation

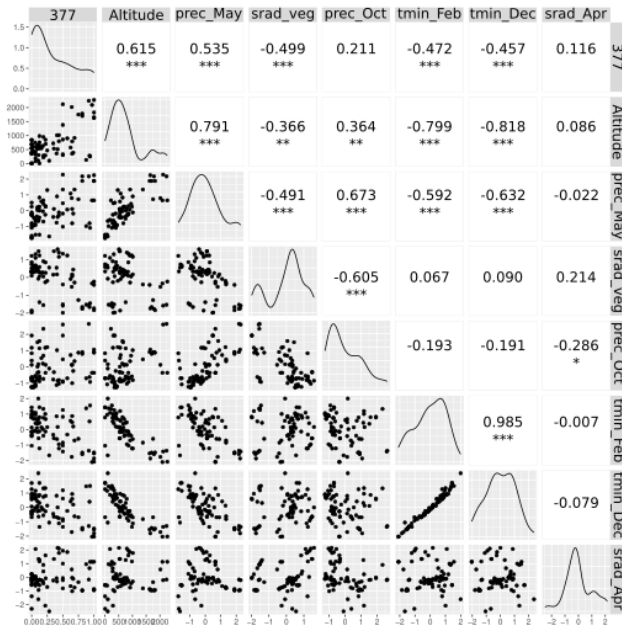
My next aim was to gain insight into the possible functional consequences of the identified *FAR5* alleles and metabolomic phenotypes. For this, I assessed which environmental factors may have shaped the distribution of the *FAR5 A* allele among populations of *A. arenosa*. To do so, I used partial redundancy analysis (pRDA). RDA is a powerful tool based on multivariate regression which allows to model linear relationships between genomic data and environmental variables. More precisely, I used a derived method called partial redundancy analysis (pRDA) that enabled me to separately assess the explained variability for different environmental variables and to account for the correlations between the variables (for details see Methods).

My dataset comprised geographic data from 73 populations across the distribution range of *A. arenosa* (Fig. 1A) and 71 climatic variables extracted for these coordinates from the WorldClim database. I further included data for all climatic variables during the most critical month of the growing season for each population, based on our field observations, which I further denote as ‘vegetative month’ variables (Suppl. Table 3). My final set encompassed 79 climatic variables, seven of which were selected by pRDA as significant for explaining the variation in the frequency of the *FAR5 A* allele (Table 1). Out of these, the best correlated explanatory variables suggest an association of the genetic variance of *FAR5* alleles with early season precipitation, irradiation during vegetative season or winter temperatures (Fig. 9).

Table 1: Explanatory variables significant based on pRDA; \*\*  $p < 0.01$  \*  $p < 0.05$ .

Variable	Cummulative R2 adjusted	F-value	p-value	correlation
prec_May	0.275927587	28.438	0.002**	0.535
srad_vegMonth	0.3413852	8.0565	0.006**	-0.499
prec_Oct	0.4605063	16.456	0.002**	0.211
tmin_Feb	0.5019368	6.7396	0.004**	-0.472
tmin_Dec	0.5727211	12.265	0.002**	-0.457
srad_Apr	0.6329272	11.989	0.004**	0.116
prec_Jul	0.6552639	5.2764	0.026*	

Associated variables



Variables for vegetative months

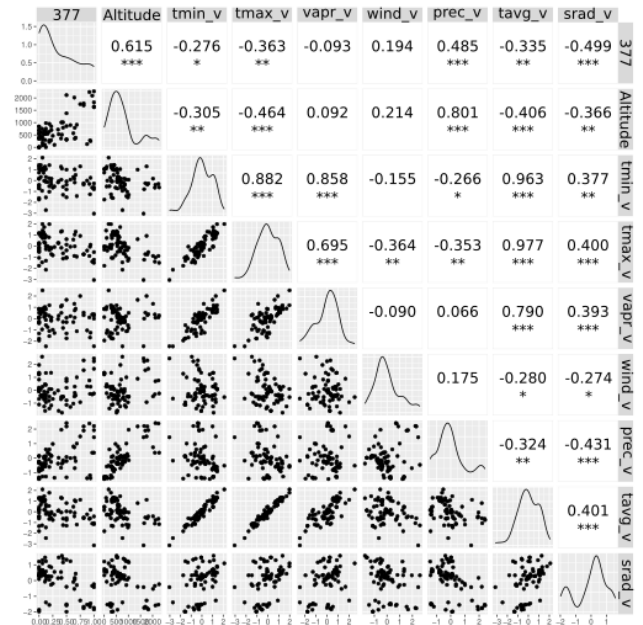


Fig. 9: Correlation of the *A* allele frequency (approximated as allele frequency at position 377) with selected environmental variables, across all 73 populations. Correlation of the best of associated variables (left) and variables for the vegetative month (right).

To examine the relationship between the best associated environmental variables and the frequency of the *A* allele, I compared their values between the populations from my dataset with almost fixed *A* allele ( $AF_A > 0.9$ ,  $N = 10$ ) and populations with fixed *F* allele ( $AF_A = 0$ ,  $N = 12$ ). I observe a significantly higher amount of precipitation in the populations with a high proportion of the *A* allele (Fig. 10BC). In the amount of solar radiation in the vegetative season and the minimum winter temperature I observe a negative correlation (Fig. 10C) and thus lower values in the populations with high proportion of the *A* allele (Fig. 10B). The relationship between the amount of precipitation in May and the frequency of the *A* allele is noticeable from the climate map (Fig. 10A), where the populations with higher proportions of the *A* allele are located mainly in the regions with higher amounts of precipitation.

To test which genes show allele frequency correspondence with the same environmental conditions as *FAR5*, and may have therefore co-evolved under a similar selective regime, I applied RDA-based GWAS. For the model I selected three climatic variables that were best associated with the frequency of the *A* allele of *FAR5* (based on the pRDA and correlation) – *prec\_May*, *srad\_vegMonth* and *tmin\_Feb*. I also accounted for the neutral genetic structure using the population scores along the first two axes of genetic PCA. I found 19 loci associated with similar environmental conditions as *FAR5*, which may have therefore evolved under similar selective forces. These 19 SNPs are annotated to 9 genes. Specifically, I detected a single SNP of a protein from the family CAMTA (calmodulin-binding transcription activators), 4 SNPs from a translation regulatory factor MRF2, 3 SNPs from *FAR5*, an arginine/serine-rich splicing factor RSP35, a transporter ABCG40 associated with ABA transport and resistance to lead, a vacuolar phosphate transporter VPT1 securing phosphate homeostasis, 3 SNPs from CFTR transporter, DOG1 controlling seed dormancy and mRNA binding CRS1 / YhbY (CRM) domain-containing protein (Suppl. Table 4). From all the SNPs, only the SNP from CAMTA and one SNP from MRF2 had stronger association with the environmental predictors than *FAR5*. The SNPs from *FAR5* were associated as follows: strongest association was with Leu377Val (based on its variability, the predictors were chosen), then Gly329Cys and lastly Ala384Ser.

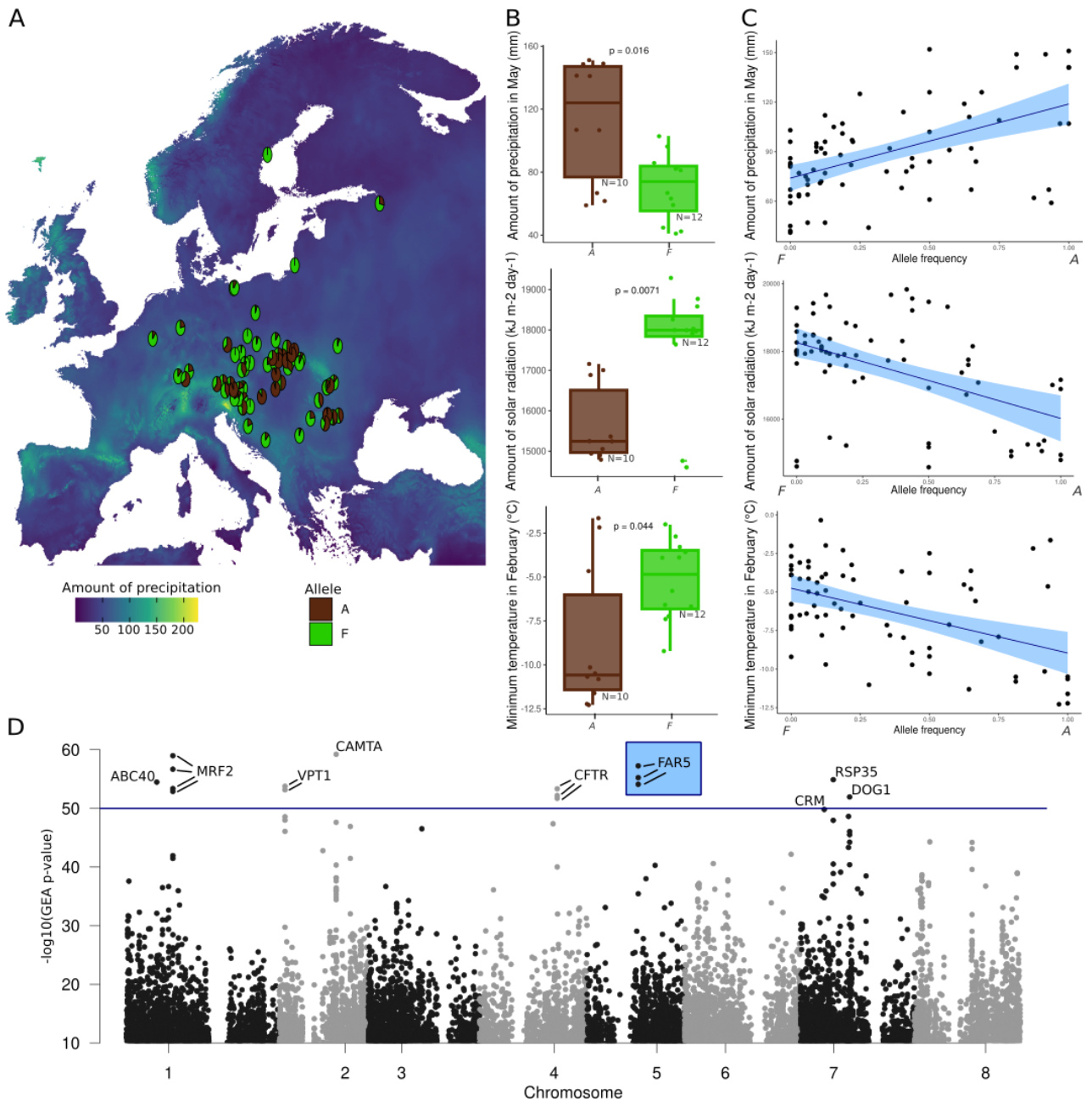


Fig. 10: Environmental drivers of FAR5 variation; A) frequency of the A and F- allele across the set of 73 populations of *A. arenosa*, projected on the map of spring precipitation (climatic variable *prec\_May*); B) Differences in three climatic variables (*prec\_May*, *srad\_vegMonth*, *tmin\_Feb*) between populations with nearly fixed A alleles ( $AFA > 0.9$ ,  $N = 10$ ) and populations with fixed F alleles ( $AFA = 0$ ,  $N = 12$ ).; C) correlation of A allele frequency with the three climatic variables (*prec\_May*, *srad\_vegMonth*, *tmin\_Feb*); D) Manhattan plot of the association between genome-wide nonsynonymous variation of *A. arenosa* and the climatic variables important for the FAR5 A-allele distribution (*prec\_May*, *srad\_vegMonth*, *tmin\_Feb*), blue line indicates the strongest outlier SNPs, rectangle marks FAR5 SNPs.

I further tested which environmental factors drive the frequency of the A allele across foothill populations in my dataset of *A. arenosa*. The motivation was to understand which environmental variables support the standing variation in the foothill populations and isolate the effect of climate on the frequency of the A allele from the effect of the alpine environment itself. The change in climate between foothill and alpine populations is prominent with strong correlations of environmental variables. Furthermore, the alpine populations are almost fixed for the A allele which could confuse the interpretation of the causative environmental variables because a change in the variable would be correlated with the increase in the A allele frequency.

From the pRDA analysis on this foothill subset, two variables resulted as significant: maximum temperature in May and water vapor pressure in vegetative season (both  $p < 0.05$ ,  $R^2$  adjusted = 0.10746/0.16982). In the correlation matrix maximum temperature in May is significant (-0.352) in contrast to water vapor pressure in vegetative season (0.243) (Fig. 11 left). From the correlation with the variables from the vegetative season, I infer a slight significance of solar radiation on the frequency of *A* allele in foothill populations (Fig. 11 right).

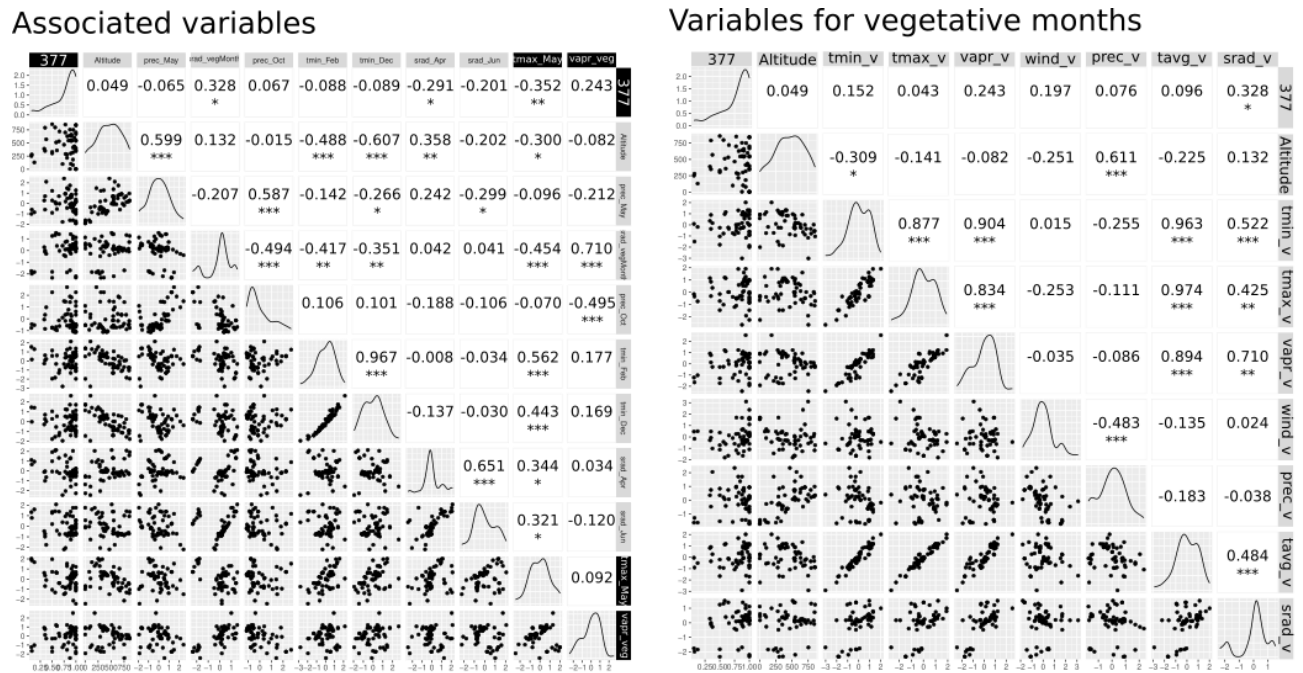


Fig. 11: Correlation of the *FAR5 A* allele frequency (approximated as allele frequency at position 377) with selected environmental variables, across the set of 55 foothill populations. Correlation of the best of associated variables (left) and variables for the representative vegetative month (right).

Despite the foothill character of the Pannonian region, there is a high frequency of the *A* allele, suggesting, perhaps a bit surprisingly, a shared selective regime on *FAR5* in this region and in the alpine region (Fig. 10A). Therefore, I also associated the explanatory variables with the subset of foothill populations without the populations located in this region. Only one environmental variable was significantly associated by the pRDA and that was the wind speed in the vegetative month ( $p = 0.018$ ,  $R^2$  adjusted = 0.09653,  $\text{corr} = 0.339$ ). Correlation between the frequency of the *A* allele and the variables for vegetative months showed also a significant correlation with the amount of precipitation (-0.305, Suppl. Fig. 10).

# Discussion

Repeated adaptation provides a unique opportunity for insight into the mechanisms of evolution. The repeated use of pathways or genes suggests their evolutionary importance. This study contributes to understanding the characteristics of a repeatedly selected allele. Thus, my main objective in this chapter is to discuss why the *FAR5 A* allele was repeatedly positively selected in the alpine environment. Here I compare results of this study with the available literature dealing with similar themes.

## *FAR5* as a candidate gene for alpine adaptation

I firstly identified *FAR5* as a strong positively selected candidate gene in alpine adaptation and characterized the specific sites and variants which contribute to this signal. The identification of *FAR5* as a strongly repeatedly selected gene is consistent with the genomic study of alpine adaptation of *A. arenosa* (Bohutínská et al., 2021b). Highly differentiated SNPs of *FAR5* between foothill and alpine populations correspond to amino acids at positions 329, 377 and 384 and to one SNP in an intron. Especially interesting is the change at the position 377 which was shown to alter *FAR5* substrate specificity in *Arabidopsis thaliana* (Chacón et al., 2013).

## The *FAR5* variation from a phylogenetic perspective

I explored the variation between my foothill and alpine samples from a broader phylogenetic perspective. From the alignment across the family *Brassicaceae* I observed that the position 384 is rather conserved. However, *FAR5 A* allele of *A. arenosa* (*A* allele further) shared the same amino acid change with *E. salsugineum*. That is an extremophile plant studied for drought and salt tolerance because of its halophytic habitat (Lugan et al., 2010). Other two candidate SNPs in *E. salsugineum* are not differentiated in the same manner as the *A* allele of *A. arenosa*. Potentially interesting could be also the change in *L. sativum* at this position because its threonine has similar chemical properties as serine present in the *A* allele. The change at the position 329 was unique in my dataset but there were two similar amino acid changes to serine in *Cakile maritima* and *Diptychocarpus strictus* and to asparagine in *Eruca vesicaria*. These two amino acids are both hydrophilic as is cysteine in the *F* allele in *A. arenosa*. There are several studies dealing with different aspects of adaptation of these species because they are inhabiting salty or sandy and arid localities. The halophytic species *C. maritima* is studied due to its potential for saline agriculture (Arbelet-Bonnin et al., 2019). The factors affecting germination were examined in *E. vesicaria* (Barazani et al., 2012) and in cold desert species *D. strictus* (Lu et al., 2010). Even though the position 329 is unique in the *F* allele the reversal to the ancestral state in the *A* allele could be functional, possibly related to water management. Interestingly, the expression patterns of *FAR5* suggest its influence on seed coat (Domergue et al., 2010; Vishwanath et al., 2013). Altogether, these phylogenetic parallels may be a first indication that the different alleles of *FAR5* could be drought-related.

From the biochemical and phylogenetic point of view, the positions 329 and 384 are more interesting than the position 377. That is surprising because of a contrast with the differentiation

pattern where the position 377 shows the strongest signal. It might suggest that all three SNPs are important for adaptive change.

The correspondence between alpine alleles of *A. arenosa* and reference alleles of *A. lyrata* at three out of four candidate positions could be related to *A. lyrata*'s occurrence at high latitude (Schmickl et al., 2010), as high latitude is known to share many characteristics with high altitude (Stevens, 1992).

## Repeated positive selection likely acted on standing variation in *FAR5*

I propose the repeated adaptation via the *A* allele of *FAR5* as a case of adaptation from standing variation. The reasoning behind this is that the candidate selected amino acid changes are present in the foothill samples as standing variation and they are all three together repeatedly found at high frequencies in all five alpine lineages. The selective sweep appears to be narrow (only 4 differentiated SNPs, <1kb apart), suggesting a longer existence of this haplotype, a signature which differentiates sign of selection acting on standing allele from the sign of selection acting on recently introgressed allele (Lee and Coop, 2019, 2017), thus giving time for recombination in its surroundings (Barrett and Schluter, 2008).

The high frequency of the *A* allele in the Pannonian lineage could be a result of incomplete lineage sorting and following action of positive selection or genetic drift in this lineage. Similar pattern was observed in some meiotic genes of *A. arenosa*, which show high differentiation in the polyploid and in the seemingly selective-regimewise unrelated diploid Pannonian lineage in the study of adaptation to polyploidy (Wright et al., 2015; Bohutínská et al., 2021a). The incomplete lineage sorting could theoretically extend as far in the phylogeny as is the divergence between *A. arenosa* and *A. lyrata* due to shared variation between *A. lyrata* and *A* allele in *A. arenosa*. To explore this possibility, a detailed study of these two species would be necessary.

## Multiple non-synonymous SNPs under selection: compensatory evolution, co-evolution or hitchhiking?

Here I show that all three identified coding SNPs differentiated between the alpine and foothill populations of *A. arenosa* are non-synonymous. The repeated positive selection of such a set of changes suggests some adaptive significance. Non-synonymous changes were reported to be adaptive before (Klim et al., 2024). The non-synonymous changes are often associated with loss of function, but they can lead to novel function of proteins (Han et al., 2022). I propose that the differentiation in *FAR5* between the foothill and alpine populations in my dataset leads to a change of function rather than a loss of function, due to the absence of coding insertions, deletions or premature stop codons. Furthermore, I elaborate on what could have contributed to the repeated positive selection of all three SNPs. I offer two possibilities: 1) only one SNP is functional and the others compensate for its possible negative side-effects or 2) all three SNPs are relevant and we witness the consequences of shared selection of all of them, possibly suggesting co-evolution or multi nucleotide mutation event. Finally, there is also a possibility of just one SNP being evolutionarily relevant and the others hitchhiking in the selective sweep.

## Compensatory evolution

The connection between different amino acids in the protein can lead to unexpected evolutionary events. The hypothesis of compensatory mutations suggests that an amino acid change affecting the structure or function of the protein can be followed by a change at a different position in an attempt to compensate for the mutation. An attempt to return the protein to the original state (DePristo et al., 2005; Kimura, 1985). It would be possible that the multiple SNPs identified in my samples are caused by FAR5's efforts to return to its original function. However, the shift in substrate specificity towards the shorter fatty alcohol could be an optimal state for the alpine environment and thus the hypothesis of compensatory mutations acting here would not be supported.

## Adaptation via multiple SNPs in one protein

Another possibility is that all the three amino acids are necessary for the shift in substrate specificity. In a non-selective scenario, they could arise during a multi-nucleotide mutation event (Schridder et al., 2011). But this scenario is not supported by the fact that in the closely related *A. thaliana* the substrate specificity changed with different amino acid changes. Moreover, I suggest that the *A* allele emerged from a standing variation in *A. arenosa* as I stated before.

Nevertheless, natural selection could act on multiple SNPs at once, resulting in their co-evolution. Several amino acid changes are often reported to affect the protein specificity or stability. Mutations at different positions can have separate effects on the enzyme specificity and stability (Dickmann et al., 2004) or a cumulative effect of multiple amino acid substitutions is necessary to change the enzyme specificity even though just one of the amino acids is in direct contact with the substrate (Oue et al., 1999). A non-synonymous substitution in cancer-associated enzyme IDH1 was previously thought to be a loss of function mutation. However, when one amino acid is changed the residues in the active site reorganize and enable the enzyme to change its substrate specificity (Dang et al., 2009). The single amino acid change is thanks to its placement in the 3D structure actually able to reposition some highly conserved amino acids at different positions of the protein.

Altogether, while we currently cannot distinguish between compensatory evolution, co-evolution or a mere hitchhiking, the presence of multiple amino acid changes selected together in the FAR5 *A-allele* may motivate further inquiries.

# FAR5 enzyme in the alpine adaptation

## The role of variation in affecting FAR5 properties

Between the *A* and *F* allele I identified four differentiated SNPs, one of them being an Ala384Ser change. This mutation induces a significant chemical change from a hydrophobic amino acid to a hydrophilic amino acid. Similar change from phenylalanine to serine was shown to affect both substrate specificity and thermal stability of cysteine sulfinic acid decarboxylase (CSAD) in mice (Mahootchi et al., 2021). However, this does not mean that the other SNPs are meaningless.

To specify if all the SNPs are truly responsible for the change in substrate specificity observed in my samples, I could construct various combinations of single and double mutants from my three identified SNPs (Cys329Gly, Val377Leu and Ala384Ser) and compare the resulting effect on enzyme behavior. This approach was previously used to successfully determine that only one of



three SNPs was responsible for altering the enzyme substrate specificity in *Paenibacillus pabuli* (Sahnoun et al., 2022).

Although directed mutagenesis is widely used in research of the enzymatic properties (Onuffer and Kirsch, 1995; Wrenbeck et al., 2017), study of the natural variation in enzymes is scarce. Palzkill, 2018 shows that natural mutations in enzymes changed the substrate specificity in order to build a resistance to antibiotics in Gram-negative bacteria, which also affected thermal stability of the protein. Nevertheless, this was compensated by the emergence of additional global suppressor mutations (Palzkill, 2018). In *Brassicaceae* the site-directed mutagenesis was used to investigate the natural variation in methylthioalkylmalate synthases (MAMs) and its effect on glucosinolate biosynthesis (Petersen et al., 2019). My study contributes to the knowledge of changes in substrate specificity and natural variation of the protein sequence of *FAR5*. By perfectly understanding the changes of protein sequence, we can achieve precise predictions about the structure and function and in turn improve the enzyme engineering (Acebes et al., 2016).

## The phenotypic effect of positively selected *FAR5* mutations

One of the main aims of this study was to determine the phenotypic manifestation of the derived *A* allele of *FAR5* gene. There was a significant difference in the proportion of C18:0-OH to C16:0-OH between plants with derived alpine allele and plants with ancestral foothill allele. Plants with the ancestral foothill allele produced more C18:0-OH than C16:0-OH whereas plants with derived alpine allele produced more C16:0-OH when compared to C18:0-OH. *FAR5* was noted to influence production of primary alcohols of different lengths in wheat leaves (C22:0-OH in functional analysis of *TaFAR5* in yeast and C26:0, C28:0 and C30:0-OH in transgenic tomato) (Wang et al., 2015), but data from the most closely related model organism, *Arabidopsis thaliana*, support the role in producing C18:0-OH (Chacón et al., 2013; Domergue et al., 2010; Vishwanath et al., 2013).

Moreover, by a series of domain swaps particular amino acids underlying substrate specificity of the enzyme were determined. One of the amino acids was at position 377 (Chacón et al., 2013). From my data I know that this position is also altered in the derived alpine allele. Apart from this amino acid also the position 355 is mentioned to affect *A. thaliana*'s substrate specificity. In particular, amino acids at both positions 355 and 377 needed to be altered to fully shift the substrate specificity of the enzyme (Chacón et al., 2013). However, my genomic data do not suggest any positively selected mutation at the position 355. It is possible that if this amino acid was also altered, the difference in fatty alcohol composition would be more striking. Another possibility is that position 377 is solely responsible for the substrate specificity of *FAR5* or that the function of the amino acid on the position 355 is complemented by mutation at position 329 or 384.

The fatty alcohols are utilized during protection against abiotic and biotic stresses. They are part of the suberin layer in roots (de Silva et al., 2021; Domergue et al., 2010; Vishwanath et al., 2013). However, the method I used to extract lipidic compounds doesn't allow us to exactly determine that the difference in composition of fatty alcohols comes from suberin. It could be from other lipidic parts of the roots, such as soluble suberin-associated waxes (Delude et al., 2016). Nevertheless, it was shown that the shorter the alcohol chain, the more likely it is to be part of suberin. (Vishwanath et al., 2013) Therefore I cannot rule out the possibility of different *FAR5* alleles affecting suberin composition.

Precise characterization of different alleles and in consequence the effect of their corresponding enzymes on fatty alcohol production may be of importance in research of sustainable wax esters which are fundamental in the production of lubricants, pharmaceuticals and cosmetics (Domergue and Miklaszewska, 2022).

# Role of environmental factors in the *FAR5*-mediated adaptation to the alpine environment

## Environmental drivers of *FAR5* variation

Abiotic environmental factors present an important obstacle in the establishment of plants in the alpine environment. Therefore, I could not omit them in my study of alpine adaptation of *Arabidopsis arenosa*. I used partial redundancy analysis to associate the genetic variation in *FAR5* with the environmental predictors. My results show that increased precipitation in the spring is an important factor, positively affecting the frequency of the *A* allele. Several other studies also reported possibly adaptive associations with precipitation in their genotype-environment association studies in Norway spruce (Di Pierro et al., 2016), two snowbed species, *Achillea clusiana* and *Campanula pulla* (Felkel et al., 2023) and a set of 13 alpine plant species (Manel et al., 2012) all conducted in the range of Alps. That is consistent with the fact that precipitation and temperature are among the major ecological variables determining the distribution of plants and driving their adaptation (Berry and Bjorkman, 1980).

The lowest winter temperature was also identified as a significant factor in the distribution of genetic variation in *FAR5*. Low temperature acts as a limiting factor for the distribution of many plants, and freezing resistance has been repeatedly investigated also in *Arabidopsis* (Hannah et al., 2006; Kaplenig et al., 2022). Minimum temperature was shown to be important for the genetic diversity of four conifer species from the European Alps (Mosca et al., 2012). In the same region, the adaptive genetic variation of 13 alpine plant species was associated with temperature as one of two major explanatory factors, the other being precipitation (Manel et al., 2012).

Another essential factor for the well-being of plants is solar radiation (Yang et al., 2022). The usual expectation is that the amount of solar radiation increases with altitude (Steinhauser et al., 1958) according to (Körner, 2022). However in my dataset, I observe a negative correlation of the *A* allele frequency and solar radiation in the vegetative season. That is arising from the fact that each population had a different month as its most important vegetative month and the populations with higher frequency of *A* allele vegetate in periods with short days. In general however, the increasing trend of solar radiation with altitude is not so prominent due to increasing cloudiness in the high elevations (Körner, 2022). In spite of that, the study of alpine adaptation in Tibetan poplar (*Populus szechuanica* var. *tibetica*) identified two hotspot regions with robust signals of natural selection showing an association with altitude and solar radiation (Zheng et al., 2020).

The changes in solar radiation are associated with a tendency of increasing proportion of ultraviolet radiation with altitude which can affect plant life severely, thus posing an obstacle to plant alpine adaptation (Caldwell, 1968). Alpine plants are often preventing damage caused by the UV radiation via increased pigmentation, altered cuticle width and composition and trichome density (Koski and Ashman, 2016, 2015). The pigments can also protect the plants against drought

or cold temperatures (Chalker-Scott, 1999). That is probably why we see genes associated with response to UV and flavonoid biosynthesis as selected outliers in some cases of alpine adaptation (Sun et al., 2020). Suberin has not been reported to be related to solar radiation. I therefore hypothesize that the association with the *FAR5 A* allele could be related indirectly, for example, through the day length-dependent timing of seed germination via seed coat composition.

When examining what other genes were associated with the alpine environment, I repeatedly encountered identification of genes involved in regulation of transcription and translation (Di Pierro et al., 2016; Novikova et al., 2023; Zheng et al., 2020) or abiotic stress response, e.g. heat shock proteins (Mosca et al., 2016, 2012). As another response to abiotic stress, plants often undergo lignification to improve their tolerance (Fan et al., 2006). Two genes involved in lignin biosynthesis were associated in the study of alpine adaptation in conifers: cinnamoyl-CoA reductase (CCR1) and pinoresinol reductase (PRR1) (Mosca et al., 2012). In alpine adaptation in Siberian larch (*Larix sibirica*) several genes were identified as candidate genes associated with altitude and other bioclimatic variables (Novikova et al., 2023). One of the candidates was *FAR4*, suggesting an unrecognized importance of FAR proteins for adaptation to the alpine environment. The production of secondary compounds, including suberin, is associated with exposure to environmental stress (Chalker-Scott, 1999), hence the relevance of FARs for alpine adaptation could lie in their ability to affect suberin layer and consequently the protection of plants against drought and cold temperatures (Gou et al., 2009). In the alpine populations of two *Pinus* species there is a signal of selection for acyl-CoA oxidase associated with annual temperature (Mosca et al., 2016). This supports the importance of lipid metabolism in alpine adaptation.

## Exploring the adaptive role of *FAR5*

After discovering a specific phenotypic effect of the *A* allele of *FAR5*, I could not help but to wonder about its functional impact on the alpine adaptation. Considering all the gathered information about phenotype, transcriptomic pattern, associated environmental variables, signs of co-evolution and available literature I believe that main reasons for repeated positive selection lie in its putative effect on seed dormancy, pathogen defense and/or on abiotic stress such as drought.

The effect on seed dormancy via specific suberin composition is supported by two genes showing association with similar environmental conditions as *FAR5* in *A. arenosa*. First is *ABC40* which is a membrane protein from the ABC family. The ABC transporters are believed to aid the suberin monomers in their transport through the membrane (Pighin et al., 2004; Rains et al., 2018). Mutants in another protein from the ABC family, *ABCG1*, demonstrated altered root suberin composition. Specifically, the *abcg1* mutants had reduced abundance of longer-chain (C20+) dicarboxylic acids, fatty alcohols and acids (Shanmugarajah et al., 2019). Furthermore, the role of *ABC40* in abscisic acid (ABA) transport could be of significance considering suberization being stimulated by ABA in potato (Cottle and Kolattukudy, 1982) and the role of ABA as a phytohormone in timing of the seed dormancy (Schopfer et al., 1979).

The second interesting gene was *DOG1* due to its involvement in the seasonal timing of germination (Huo et al., 2016). Since *FAR5* is expressed in the seed coat of *A. thaliana* (Domergue et al., 2010), it could also affect germination. Moreover, suberin has been shown as a protective factor for seed germination in the presence of chromium (Cr<sup>3+</sup>) (de Silva et al., 2021).

In a genomic study of pathogen response in *A. thaliana*, *FAR5* was identified as one of three significant candidates (Kirischian, 2017). This suggests a yet unknown role in plant immunity of FAR proteins suggested already by (Domergue et al., 2010) due to their effect on suberin which

serves as a barrier for pathogens. It is also supported by antibacterial properties of long chain fatty alcohols (Hattori et al., 1987).

The production of suberin was previously associated with plant reactions to abiotic stress, such as cold, salt stress (Gou et al., 2009) or drought (Franke et al., 2012). Therefore, the ability to appropriately modify it could be beneficial during the adaptation to the alpine environment, whether for seeds or mature plants.

The apparently specific effect of the *A* allele on the production of C16:0-OH/C18:0-OH may only be only one piece of the range of the *A* allele's impact on local adaptation given the possible importance of *FAR5* e.g. in response to pathogens. Larger phenotypic effect size of a candidate adaptive gene would correspond to a preprint showing that repeatedly selected genes in local adaptation to climate variation have higher pleiotropy than originally thought (Yeaman et al., 2023).

Overall, the complexity of the alpine environment hinders the efforts to describe the mechanisms of local adaptation accurately. There are many biotic and abiotic factors affecting the strenuous adaptation of organisms resulting in polygenic character of the adaptation. Nevertheless, I believe that this and other future studies will significantly contribute to the understanding of the big picture of alpine adaptation.

# Conclusions

In this study, I present my findings regarding the role of the candidate alpine-adaptive *A* allele of *FAR5* in the alpine adaptation of *Arabidopsis arenosa*. I revealed a detectable phenotype of the candidate allele and observed a signal of directed change in the substrate specificity in the alpine environment.

Firstly, I established the importance of *FAR5* in alpine adaptation by showing that it was highly differentiated between the foothill and alpine environments throughout the five alpine colonization events. I characterized the two alleles, *F* allele typical for foothill populations and *A* allele typical for alpine populations. I showed three linked coding SNPs and one intron SNPs as differentiating the two alleles.

Subsequently, I showed that plants with the *A* allele produced more C16:0-OH than plants with the *F* allele. The latter produced more C18:0-OH. This pattern was consistent across all my samples in both measured tissues, roots and wounded leaves. These results suggest a change in substrate specificity between the variants of *FAR5* corresponding to the two alleles.

Finally, I associated the genetic variation in *FAR5* with the environmental factors. I observed the strongest association between allele frequencies and early season precipitation. Among other strong factors were solar radiation during the vegetative season and minimum winter temperature. All these factors can be associated with the challenges the alpine environment presents for plant survival. Furthermore, I identified eight genes selected under similar environmental conditions as *FAR5*. The genes are involved in timing of germination, regulation of transcription and translation and transport.

To increase our knowledge about the specific function of the enzyme and the importance of observed metabolic changes for plant life in the alpine environment, future research should be devoted to the examination of the seed coat, the effect on germination and transcriptomic analysis.

The uniqueness of this study lies in the functional validation of a clearly defined candidate allele. That is quite rare in evolutionary biology. A practical use might come from the broad industrial application of enzymes, for *FAR5* particularly in pharmaceuticals and cosmetics as the produced fatty alcohols are fundamental for formation of wax esters utilized in production of bio-lubricants. From the evolutionary aspect, this study contributes to the knowledge about adaptation to the steep change of conditions between the foothill and alpine environment. Understanding the molecular mechanisms of such adaptation could be beneficial in the future in the light of recent environmental changes.

# Bibliography

- Acebes, S., Fernandez-Fueyo, E., Monza, E., Lucas, M.F., Almendral, D., Ruiz-Dueñas, F.J., Lund, H., Martinez, A.T., Guallar, V., 2016. Rational Enzyme Engineering Through Biophysical and Biochemical Modeling. *ACS Catal.* 6, 1624–1629.  
<https://doi.org/10.1021/acscatal.6b00028>
- Alexander, D.E., 2015. *On the Wing*. Oxford University Press, USA.
- Arbelet-Bonnin, D., Ben-Hamed-Louati, I., Laurenti, P., Abdelly, C., Ben-Hamed, K., Bouteau, F., 2019. *Cakile maritima*, a promising model for halophyte studies and a putative cash crop for saline agriculture, in: Sparks, D.L. (Ed.), *Advances in Agronomy*. Academic Press, pp. 45–78. <https://doi.org/10.1016/bs.agron.2019.01.003>
- Arendt, J., Reznick, D., 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* 23, 26–32.  
<https://doi.org/10.1016/j.tree.2007.09.011>
- Barazani, O., Quaye, M., Ohali, S., Barzilai, M., Kigel, J., 2012. Photo-thermal regulation of seed germination in natural populations of *Eruca sativa* Miller (Brassicaceae). *J. Arid Environ.* 85, 93–96. <https://doi.org/10.1016/j.jaridenv.2012.06.011>
- Barrett, R.D.H., Schluter, D., 2008. Adaptation from standing genetic variation. *Trends Ecol. Evol.* 23, 38–44. <https://doi.org/10.1016/j.tree.2007.09.008>
- Bartok, A., Hurdu, B.-I., Szatmari, P.-M., Ronikier, M., Puşcaş, M., Novikov, A., Bartha, L., Vonica, G., 2016. New records for the high-mountain flora of the fâgăraş mts. (southern carpathians) with discussion on ecological preferences and distribution of studied taxa in the carpathians. *Contrib. Bot.* 51, 77–153.
- Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., Huala, E., 2015. The arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. *genesis* 53, 474–485. <https://doi.org/10.1002/dvg.22877>
- Bernards, M.A., Lopez, M.L., Zajicek, J., Lewis, N.G., 1995. Hydroxycinnamic Acid-derived Polymers Constitute the Polyaromatic Domain of Suberin (\*). *J. Biol. Chem.* 270, 7382–7386. <https://doi.org/10.1074/jbc.270.13.7382>
- Berner, D., 2019. Allele Frequency Difference AFD—An Intuitive Alternative to FST for Quantifying Genetic Population Differentiation. *Genes* 10, 308.  
<https://doi.org/10.3390/genes10040308>
- Berry, J., Bjorkman, O., 1980. Photosynthetic Response and Adaptation to Temperature in Higher Plants. *Annu. Rev. Plant Biol.* 31, 491–543.  
<https://doi.org/10.1146/annurev.pp.31.060180.002423>
- Birkeland, S., Gustafsson, A.L.S., Brysting, A.K., Brochmann, C., Nowak, M.D., Purugganan, M., 2020. Multiple Genetic Trajectories to Extreme Abiotic Stress Adaptation in Arctic Brassicaceae. *Mol. Biol. Evol.* 37, 2052–2068.  
<https://doi.org/10.1093/MOLBEV/MSAA068>
- Blanquart, F., Kaltz, O., Nuismer, S.L., Gandon, S., 2013. A practical guide to measuring local adaptation. *Ecol. Lett.* 16, 1195–1205. <https://doi.org/10.1111/ele.12150>
- Bohutínská, M., Handrick, V., Yant, L., Schmickl, R., Kolář, F., Bomblies, K., Paajanen, P., 2021a. De Novo Mutation and Rapid Protein (Co-)evolution during Meiotic Adaptation in *Arabidopsis arenosa*. *Mol. Biol. Evol.* 38, 1980–1994.  
<https://doi.org/10.1093/molbev/msab001>

- Bohutínská, M., Vlček, J., Yair, S., Laenen, B., Konečná, V., Fracassetti, M., Slotte, T., Kolář, F., 2021b. Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives. *PNAS* 118.  
[https://doi.org/10.1073/PNAS.2022713118/SUPPL\\_FILE/PNAS.2022713118.SD10.TXT](https://doi.org/10.1073/PNAS.2022713118/SUPPL_FILE/PNAS.2022713118.SD10.TXT)
- Bombliès, K., Peichel, C.L., 2022. Genetics of adaptation. *Proc. Natl. Acad. Sci.* 119, e2122152119. <https://doi.org/10.1073/PNAS.2122152119>
- Booker, T.R., Yeaman, S., Whitlock, M.C., 2023. Using genome scans to identify genes used repeatedly for adaptation. *Evol. Int. J. Org. Evol.* 77, 801–811.  
<https://doi.org/10.1093/evolut/qpac063>
- Caldwell, M.M., 1968. Solar Ultraviolet Radiation as an Ecological Factor for Alpine Plants. *Ecol. Monogr.* 38, 243–268. <https://doi.org/10.2307/1942430>
- Capblancq, T., Forester, B.R., 2021. Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods Ecol. Evol.* 12, 2298–2309. <https://doi.org/10.1111/2041-210X.13722>
- Ceppi, P., Scherrer, S.C., Fischer, A.M., Appenzeller, C., 2012. Revisiting Swiss temperature trends 1959–2008. *Int. J. Climatol.* 32, 203–213. <https://doi.org/10.1002/joc.2260>
- Cerca, J., 2023. Understanding natural selection and similarity: Convergent, parallel and repeated evolution. *Mol. Ecol.* 32, 5451–5462. <https://doi.org/10.1111/mec.17132>
- Chacón, M.G., Fournier, A.E., Tran, F., Dittrich-Domergue, F., Pulsifer, I.P., Domergue, F., Rowland, O., 2013. Identification of amino acids conferring chain length substrate specificities on fatty alcohol-forming reductases FAR5 and FAR8 from *Arabidopsis thaliana*. *J. Biol. Chem.* 288, 30345–30355. <https://doi.org/10.1074/jbc.M113.499715>
- Chalker-Scott, L., 1999. Environmental Significance of Anthocyanins in Plant Stress Responses. *Photochem. Photobiol.* 70, 1–9. <https://doi.org/10.1111/j.1751-1097.1999.tb01944.x>
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* 6, 80–92. <https://doi.org/10.4161/fly.19695>
- Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Schluter, D., Kingsley, D.M., 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307, 1928–1933. <https://doi.org/10.1126/science.1107239>
- Conte, G.L., Arnegard, M.E., Peichel, C.L., Schluter, D., 2012. The probability of genetic parallelism and convergence in natural populations. *Proc. R. Soc. B Biol. Sci.* 279, 5039–5047. <https://doi.org/10.1098/rspb.2012.2146>
- Cottle, W., Kolattukudy, P.E., 1982. Abscisic Acid stimulation of suberization : induction of enzymes and deposition of polymeric components and associated waxes in tissue cultures of potato tuber. *Plant Physiol.* 70, 775–780. <https://doi.org/10.1104/pp.70.3.775>
- Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., Marks, K.M., Prins, R.M., Ward, P.S., Yen, K.E., Liao, L.M., Rabinowitz, J.D., Cantley, L.C., Thompson, C.B., Vander Heiden, M.G., Su, S.M., 2009. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 739. <https://doi.org/10.1038/nature08617>
- de Silva, N.D.G., Boutin, C., Lukina, A.O., Western, T.L., Molina, I., Rowland, O., 2021. Seed coat suberin forms a barrier against chromium (Cr<sup>3+</sup>) during early seed germination in *Arabidopsis thaliana*. *Environ. Exp. Bot.* 191, 104632.  
<https://doi.org/10.1016/j.envexpbot.2021.104632>
- Dean, B.B., Kolattukudy, P.E., 1976. Synthesis of Suberin during Wound-healing in Jade Leaves,

- Tomato Fruit, and Bean Pods 1. *Plant Physiol.* 58, 411–416.
- Delude, C., Fouillen, L., Bhar, P., Cardinal, M.-J., Pascal, S., Santos, P., Kosma, D.K., Joubès, J., Rowland, O., Domergue, F., 2016. Primary Fatty Alcohols Are Major Components of Suberized Root Tissues of *Arabidopsis* in the Form of Alkyl Hydroxycinnamates1[OPEN]. *Plant Physiol.* 171, 1934–1950. <https://doi.org/10.1104/pp.16.00834>
- Delude, C., Vishwanath, S., Rowland, O., Domergue, F., 2017. Root Aliphatic Suberin Analysis Using Non-extraction or Solvent-extraction Methods. *BIO-Protoc.* 7. <https://doi.org/10.21769/BioProtoc.2331>
- DePristo, M.A., Weinreich, D.M., Hartl, D.L., 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* 6, 678–687. <https://doi.org/10.1038/nrg1672>
- Di Pierro, E.A., Mosca, E., Rocchini, D., Binelli, G., Neale, D.B., La Porta, N., 2016. Climate-related adaptive genetic variation and population structure in natural stands of Norway spruce in the South-Eastern Alps. *Tree Genet. Genomes* 12, 16. <https://doi.org/10.1007/s11295-016-0972-4>
- Dickmann, L.J., Locuson, C.W., Jones, J.P., Rettie, A.E., 2004. Differential Roles of Arg97, Asp293, and Arg108 in Enzyme Stability and Substrate Specificity of CYP2C9. *Mol. Pharmacol.* 65, 842–850. <https://doi.org/10.1124/mol.65.4.842>
- Domergue, F., Miklaszewska, M., 2022. The production of wax esters in transgenic plants: towards a sustainable source of bio-lubricants. *J. Exp. Bot.* 73, 2817–2834. <https://doi.org/10.1093/jxb/erac046>
- Domergue, F., Vishwanath, S.J., Joubès, J., Ono, J., Lee, J.A., Bourdon, M., Alhattab, R., Lowe, C., Pascal, S., Lessire, R., Rowland, O., 2010. Three *Arabidopsis* fatty acyl-coenzyme A reductases, FAR1, FAR4, and FAR5, generate primary fatty alcohols associated with suberin deposition. *Plant Physiol.* 153, 1539–1554. <https://doi.org/10.1104/PP.110.158238>
- Fan, L., Linker, R., Gepstein, S., Tanimoto, E., Yamamoto, R., Neumann, P.M., 2006. Progressive Inhibition by Water Deficit of Cell Wall Extensibility and Growth along the Elongation Zone of Maize Roots Is Related to Increased Lignin Metabolism and Progressive Stelar Accumulation of Wall Phenolics. *Plant Physiol.* 140, 603–612. <https://doi.org/10.1104/pp.105.073130>
- Felkel, S., Tremetsberger, K., Moser, D., Dohm, J.C., Himmelbauer, H., Winkler, M., 2023. Genome-environment associations along elevation gradients in two snowbed species of the North-Eastern Calcareous Alps. *BMC Plant Biol.* 23, 203. <https://doi.org/10.1186/s12870-023-04187-x>
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. <https://doi.org/10.1002/joc.5086>
- Flohn, H., 1974. Contribution to a comparative meteorology of mountain areas, in: *Arctic and Alpine Environments*. Methuen, London, p. pp 55-71.
- Franke, R., Schreiber, L., 2007. Suberin — a biopolyester forming apoplastic plant interfaces. *Curr. Opin. Plant Biol., Physiology and Metabolism* 10, 252–259. <https://doi.org/10.1016/j.pbi.2007.04.004>
- Franke, R.B., Dombrink, I., Schreiber, L., 2012. Suberin Goes Genomics: Use of a Short Living Plant to Investigate a Long Lasting Polymer. *Front. Plant Sci.* 3. <https://doi.org/10.3389/fpls.2012.00004>
- Fraser, B.A., Whiting, J.R., 2020. What can be learned by scanning the genome for molecular convergence in wild populations? *Ann. N. Y. Acad. Sci.* 1476, 23–42.



<https://doi.org/10.1111/NYAS.14177>

- Gale, J., 1972. Availability of Carbon Dioxide for Photosynthesis at High Altitudes: Theoretical Considerations. *Ecology* 53, 494–497. <https://doi.org/10.2307/1934239>
- Gebhardt, C., 2007. Molecular Markers, Maps and Population Genetics, in: *Potato Biology and Biotechnology : Advances and Perspectives*. Elsevier Science.
- Gou, J.-Y., Yu, X.-H., Liu, C.-J., 2009. A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 106, 18855–18860. <https://doi.org/10.1073/pnas.0905555106>
- Han, H., Xu, M., Wen, L., Chen, J., Liu, Q., Wang, J., Li, M.D., Yang, Z., 2022. Identification of a Novel Functional Non-synonymous Single Nucleotide Polymorphism in Frizzled Class Receptor 6 Gene for Involvement in Depressive Symptoms. *Front. Mol. Neurosci.* 15. <https://doi.org/10.3389/fnmol.2022.882396>
- Hannah, M.A., Wiese, D., Freund, S., Fiehn, O., Heyer, A.G., Hinch, D.K., 2006. Natural Genetic Variation of Freezing Tolerance in *Arabidopsis*. *Plant Physiol.* 142, 98–112. <https://doi.org/10.1104/pp.106.081141>
- Hattori, M., Miyachi, K., Hada, S., Kakiuchi, N., Kiuchi, F., Tsuda, Y., Namba, T., 1987. Effects of Long-Chain Fatty Acids and Fatty Alcohols on the Growth of *Streptococcus mutans*. *Chem. Pharm. Bull. (Tokyo)* 35, 3507–3510. <https://doi.org/10.1248/cpb.35.3507>
- Hernangómez, D., 2024. giscoR: Download Map Data from GISCO API - Eurostat. <https://doi.org/10.5281/zenodo.4317946>
- Hijmans, R.J., 2023a. terra: Spatial Data Analysis.
- Hijmans, R.J., 2023b. raster: Geographic Data Analysis and Modeling.
- Hollister, J., Shah, T., Nowosad, J., Robitaille, A.L., Beck, M.W., Johnson, M., 2023. elevatr: Access Elevation Data from Various APIs. <https://doi.org/10.5281/zenodo.8335450>
- Howard, K.A., Card, C., Benner, J.S., Callahan, H.L., Maunus, R., Silber, K., Wilson, G., Brooks, J.E., 1986. Cloning the DdeI restriction-modification system using a two-step method. *Nucleic Acids Res.* 14, 7939–7951.
- Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.-F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J.D., Ossowski, S., Ottillar, R.P., Salamov, A.A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M.E., Bergelson, J., Carrington, J.C., Gaut, B.S., Schmutz, J., Mayer, K.F.X., Van de Peer, Y., Grigoriev, I.V., Nordborg, M., Weigel, D., Guo, Y.-L., 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* 43, 476–481. <https://doi.org/10.1038/ng.807>
- Hudson, R.R., Slatkin, M., Maddison, W.P., 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132, 583–589. <https://doi.org/10.1093/genetics/132.2.583>
- Huo, H., Wei, S., Bradford, K.J., 2016. DELAY OF GERMINATION1 (DOG1) regulates both seed dormancy and flowering time through microRNA pathways. *Proc. Natl. Acad. Sci.* 113, E2199–E2206. <https://doi.org/10.1073/pnas.1600558113>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohli, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>

- Kaplenig, D., Bertel, C., Arc, E., Villscheider, R., Ralser, M., Kolář, F., Wos, G., Hülber, K., Kranner, I., Neuner, G., 2022. Repeated colonization of alpine habitats by *Arabidopsis arenosa* viewed through freezing resistance and ice management strategies. *Plant Biol.* 24, 939–949. <https://doi.org/10.1111/plb.13454>
- Kavanagh, K.L., Jörnvall, H., Persson, B., Oppermann, U., 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cell. Mol. Life Sci.* 65, 3895. <https://doi.org/10.1007/s00018-008-8588-y>
- Kimura, M., 1985. The role of compensatory neutral mutations in molecular evolution. *J. Genet.* 64, 7–19. <https://doi.org/10.1007/BF02923549>
- Kirischian, N., 2017. Identification of novel resistance specificity in *Arabidopsis thaliana* against *Pseudomonas syringae* strains (Thesis).
- Klepikova, A.V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D., Penin, A.A., 2016. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* 88, 1058–1070. <https://doi.org/10.1111/tpj.13312>
- Klim, J., Zielenkiewicz, U., Kaczanowski, S., 2024. Loss-of-function mutations are main drivers of adaptations during short-term evolution. *Sci. Rep.* 14, 7128. <https://doi.org/10.1038/s41598-024-57694-8>
- Knotek, A., Konečná, V., Wos, G., Požárová, D., Šrámková, G., Bohutínská, M., Zeisek, V., Marhold, K., Kolář, F., 2020. Parallel Alpine Differentiation in *Arabidopsis arenosa*. *Front. Plant Sci.* 11, 1–12. <https://doi.org/10.3389/fpls.2020.561526>
- Kolář, F., Fuxová, G., Závěská, E., Nagano, A.J., Hyklová, L., Lučanová, M., Kudoh, H., Marhold, K., 2016. Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*. *Mol. Ecol.* 25, 3929–3949. <https://doi.org/10.1111/mec.13721>
- Kolattukudy, P.E., 2001. Polyesters in Higher Plants, in: Babel, W., Steinbüchel, A. (Eds.), *Biopolyesters*. Springer, Berlin, Heidelberg, pp. 1–49. [https://doi.org/10.1007/3-540-40021-4\\_1](https://doi.org/10.1007/3-540-40021-4_1)
- Konečná, V., Bray, S., Vlček, J., Bohutínská, M., Požárová, D., Choudhury, R.R., Bollmann-Giolai, A., Flis, P., Salt, D.E., Parisod, C., Yant, L., Kolář, F., 2021. Parallel adaptation in autopolyploid *Arabidopsis arenosa* is dominated by repeated recruitment of shared alleles. *Nat. Commun.* 12. <https://doi.org/10.1038/S41467-021-25256-5>
- Konieczny, A., Ausubel, F.M., 1993. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* 4, 403–410. <https://doi.org/10.1046/j.1365-313X.1993.04020403.x>
- Körner, C., 2023. Concepts in Alpine Plant Ecology. *Plants* 12, 2666. <https://doi.org/10.3390/plants12142666>
- Körner, C., 2022. *Alpine plant life*. Springer.
- Körner, Ch., Paulsen, J., Pelaez-Riedl, S., 2003. A Bioclimatic Characterisation of Europe's Alpine Areas, in: Nagy, L., Grabherr, G., Körner, Christian, Thompson, D.B.A. (Eds.), *Alpine Biodiversity in Europe*. Springer, Berlin, Heidelberg, pp. 13–28. [https://doi.org/10.1007/978-3-642-18967-8\\_2](https://doi.org/10.1007/978-3-642-18967-8_2)
- Koski, M.H., Ashman, T.-L., 2016. Macroevolutionary patterns of ultraviolet floral pigmentation explained by geography and associated bioclimatic factors. *New Phytol.* 211, 708–718. <https://doi.org/10.1111/nph.13921>
- Koski, M.H., Ashman, T.-L., 2015. An altitudinal cline in UV floral pattern corresponds with a behavioral change of a generalist pollinator assemblage. *Ecology* 96, 3343–3353.

<https://doi.org/10.1890/15-0242.1>

- Kunka, A., Marques, S.M., Havlasek, M., Vasina, M., Velatova, N., Cengelova, L., Kovar, D., Damborsky, J., Marek, M., Bednar, D., Prokop, Z., 2023. Advancing Enzyme's Stability and Catalytic Efficiency through Synergy of Force-Field Calculations, Evolutionary Analysis, and Machine Learning. *ACS Catal.* 13, 12506–12518.  
<https://doi.org/10.1021/acscatal.3c02575>
- Lee, K.M., Coop, G., 2019. Population genomics perspectives on convergent adaptation. *Philos. Trans. R. Soc. B* 374. <https://doi.org/10.1098/RSTB.2018.0236>
- Lee, K.M., Coop, G., 2017. Distinguishing Among Modes of Convergent Adaptation Using Population Genomic Data. *Genetics* 207, 1591–1619.  
<https://doi.org/10.1534/genetics.117.300417>
- Losos, J.B., 2011. Convergence, adaptation, and constraint. *Evolution* 65, 1827–1840.  
<https://doi.org/10.1111/j.1558-5646.2011.01289.x>
- Lu, J., Tan, D., Baskin, J.M., Baskin, C.C., 2010. Fruit and seed heteromorphism in the cold desert annual ephemeral *Diptychocarpus strictus* (Brassicaceae) and possible adaptive significance. *Ann. Bot.* 105, 999–1014. <https://doi.org/10.1093/aob/mcq041>
- Lugan, R., Niogret, M.-F., Lepout, L., Guégan, J.-P., Larher, F.R., Savouré, A., Kopka, J., Bouchereau, A., 2010. Metabolome and water homeostasis analysis of *Thellungiella salsuginea* suggests that dehydration tolerance is a key response to osmotic stress in this halophyte. *Plant J.* 64, 215–229. <https://doi.org/10.1111/j.1365-313X.2010.04323.x>
- Mahootchi, E., Raasakka, A., Luan, W., Muruganandam, G., Loris, R., Haavik, J., Kursula, P., 2021. Structure and substrate specificity determinants of the taurine biosynthetic enzyme cysteine sulphinic acid decarboxylase. *J. Struct. Biol.* 213, 107674.  
<https://doi.org/10.1016/j.jsb.2020.107674>
- Manel, S., Gugerli, F., Thuiller, W., Alvarez, N., Legendre, P., Holderegger, R., Gielly, L., Taberlet, P., Consortium, I., 2012. Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Mol. Ecol.* 21, 3729–3738.  
<https://doi.org/10.1111/j.1365-294X.2012.05656.x>
- Marburger, S., Monnahan, P., Seear, P.J., Martin, S.H., Koch, J., Paajanen, P., Bohutínská, M., Higgins, J.D., Schmickl, R., Yant, L., 2019. Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* 10, 1–11. <https://doi.org/10.1038/s41467-019-13159-5>
- Melzer, H., 1960. Neues und Kritisches zur Flora der Steiermark und des angrenzenden Burgenlandes.
- Mergner, J., Frejno, M., List, M., Papacek, M., Chen, X., Chaudhary, A., Samaras, P., Richter, S., Shikata, H., Messerer, M., Lang, D., Altmann, S., Cyprys, P., Zolg, D.P., Mathieson, T., Bantscheff, M., Hazarika, R.R., Schmidt, T., Dawid, C., Dunkel, A., Hofmann, T., Sprunck, S., Falter-Braun, P., Johannes, F., Mayer, K.F.X., Jürgens, G., Wilhelm, M., Baumbach, J., Grill, E., Schneitz, K., Schwechheimer, C., Kuster, B., 2020. Mass-spectrometry-based draft of the *Arabidopsis* proteome. *Nature* 579, 409–414. <https://doi.org/10.1038/s41586-020-2094-2>
- Měsíček, J., Goliášová, K., 2002. “*Cardaminopsis* (C. A. Mey.) Hayek,” in: *Flóra Slovenska*. Veda, Bratislava.
- Monnahan, P., Kolář, F., Baduel, P., Sailer, C., Koch, J., Horvath, R., Laenen, B., Schmickl, R., Paajanen, P., Šrámková, G., Bohutínská, M., Arnold, B., Weisman, C.M., Marhold, K., Slotte, T., Bomblies, K., Yant, L., 2019. Pervasive population genomic consequences of

- genome duplication in *Arabidopsis arenosa*. *Nat. Ecol. Evol.* 3, 457–468.  
<https://doi.org/10.1038/s41559-019-0807-4>
- Mosca, E., Eckert, A.J., Di Pierro, E.A., Rocchini, D., La Porta, N., Belletti, P., Neale, D.B., 2012. The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Mol. Ecol.* 21, 5530–5545.  
<https://doi.org/10.1111/mec.12043>
- Mosca, E., Gugerli, F., Eckert, A.J., Neale, D.B., 2016. Signatures of natural selection on *Pinus cembra* and *P. mugo* along elevational gradients in the Alps. *Tree Genet. Genomes* 12, 9.  
<https://doi.org/10.1007/s11295-015-0964-9>
- Novikova, P.Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., Muir, G., Guggisberg, A., Paape, T., Schmid, K., Fedorenko, O.M., Holm, S., Säll, T., Schlötterer, C., Marhold, K., Widmer, A., Sese, J., Shimizu, K.K., Weigel, D., Krämer, U., Koch, M.A., Nordborg, M., 2016. Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nat. Genet.* 48, 1077–1082.  
<https://doi.org/10.1038/ng.3617>
- Novikova, S.V., Sharov, V.V., Oreshkova, N.V., Simonov, E.P., Krutovsky, K.V., 2023. Genetic Adaptation of Siberian Larch (*Larix sibirica* Ledeb.) to High Altitudes. *Int. J. Mol. Sci.* 24, 4530. <https://doi.org/10.3390/ijms24054530>
- Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O’Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M.D., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.-H., Cunha, E.R., Smith, T., Stier, A., Braak, C.J.F.T., Weedon, J., 2022. *vegan: Community Ecology Package*.
- Onuffer, J.J., Kirsch, J.F., 1995. Redesign of the substrate specificity of *escherichia coli* aspartate aminotransferase to that of *escherichia coli* tyrosine aminotransferase by homology modeling and site-directed mutagenesis. *Protein Sci.* 4, 1750–1757.  
<https://doi.org/10.1002/pro.5560040910>
- Oue, S., Okamoto, A., Yano, T., Kagamiyama, H., 1999. Redesigning the Substrate Specificity of an Enzyme by Cumulative Effects of the Mutations of Non-active Site Residues\*. *J. Biol. Chem.* 274, 2344–2349. <https://doi.org/10.1074/jbc.274.4.2344>
- Pachschwöll, C., Pachschwöll, T., 2019. A new find of *Arabidopsis neglecta* ( Brassicaceae) in the Svydovets Massif (Ukrainian Carpathians). *Ukr. Bot. J.* 76, 60–66.
- Palzkill, T., 2018. Structural and Mechanistic Basis for Extended-Spectrum Drug-Resistance Mutations in Altering the Specificity of TEM, CTX-M, and KPC  $\beta$ -lactamases. *Front. Mol. Biosci.* 5. <https://doi.org/10.3389/fmolb.2018.00016>
- Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *R J.* 10, 439–446. <https://doi.org/10.32614/RJ-2018-009>
- Petersen, A., Hansen, L.G., Mirza, N., Crocoll, C., Mirza, O., Halkier, B.A., 2019. Changing substrate specificity and iteration of amino acid chain elongation in glucosinolate biosynthesis through targeted mutagenesis of *Arabidopsis* methylthioalkylmalate synthase 1. *Biosci. Rep.* 39, BSR20190446. <https://doi.org/10.1042/BSR20190446>
- Pighin, J.A., Zheng, H., Balakshin, L.J., Goodman, I.P., Western, T.L., Jetter, R., Kunst, L., Samuels, A.L., 2004. Plant Cuticular Lipid Export Requires an ABC Transporter. *Science* 306, 702–704. <https://doi.org/10.1126/science.1102331>
- Pollard, M., Beisson, F., Li, Y., Ohlrogge, J.B., 2008. Building lipid barriers: biosynthesis of cutin

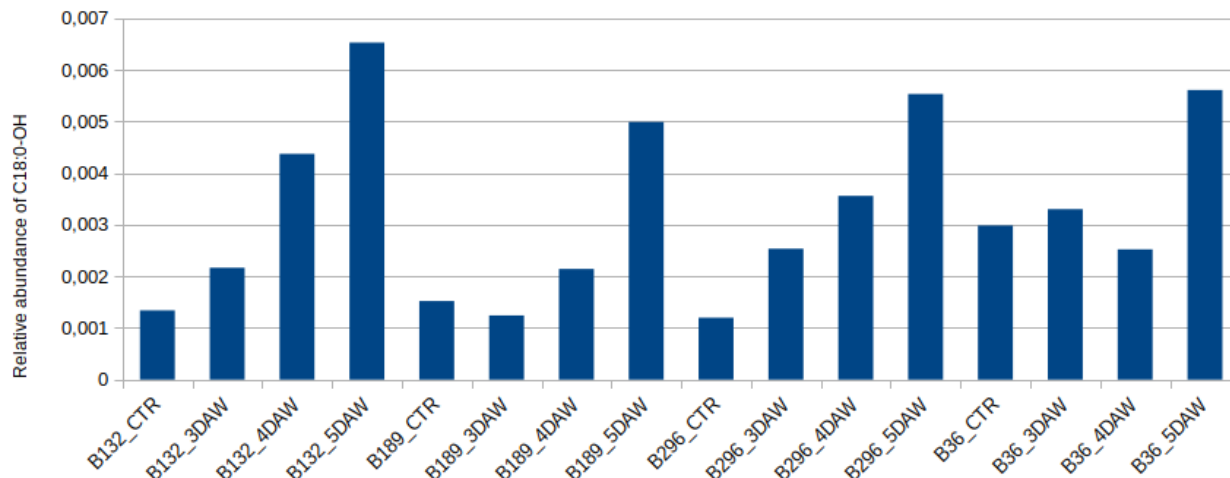
- and suberin. *Trends Plant Sci.* 13, 236–246. <https://doi.org/10.1016/j.tplants.2008.03.003>
- Preite, V., Sailer, C., Syllwasschy, L., Bray, S., Ahmadi, H., Kraemer, U., Yant, L., 2019. Convergent evolution in *Arabidopsis halleri* and *Arabidopsis arenosa* on calamine metalliferous soils. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 374, 20180243. <https://doi.org/10.1098/rstb.2018.0243>
- R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rains, M.K., Gardiyehewa de Silva, N.D., Molina, I., 2018. Reconstructing the suberin pathway in poplar by chemical and transcriptomic analysis of bark tissues. *Tree Physiol.* 38, 340–361. <https://doi.org/10.1093/treephys/tpx060>
- Rawat, V., Abdelsamad, A., Pietzenuk, B., Seymour, D.K., Koenig, D., Weigel, D., Pecinka, A., Schneeberger, K., 2015. Improving the Annotation of *Arabidopsis lyrata* Using RNA-Seq Data. *PLoS ONE* 10, e0137391. <https://doi.org/10.1371/journal.pone.0137391>
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rowland, O., Domergue, F., 2012. Plant fatty acyl reductases: Enzymes generating fatty alcohols for protective layers with potential for industrial applications. *Plant Sci.* 193–194, 28–38. <https://doi.org/10.1016/j.plantsci.2012.05.002>
- Sahnoun, M., Jaoua, M., Bejar, S., Jemli, S., 2022. Highlight on mutations affecting the US132 cyclodextrin glucanotransferase binding specificity, thermal stability, and anti-staling activity. *Colloids Surf. B Biointerfaces* 212, 112375. <https://doi.org/10.1016/j.colsurfb.2022.112375>
- Salisbury, F.B., Spomer, G.G., 1964. Leaf temperatures of alpine plants in the field. *Planta* 60, 497–505. <https://doi.org/10.1007/BF01894807>
- Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., Crowley, J., 2021. GGally: Extension to “ggplot2.”
- Schmickl, R., Jørgensen, M.H., Brysting, A.K., Koch, M.A., 2010. The evolutionary history of the *Arabidopsis lyrata* complex: a hybrid in the amphi-Beringian area closes a large distribution gap and builds up a genetic barrier. *BMC Evol. Biol.* 10, 98. <https://doi.org/10.1186/1471-2148-10-98>
- Schopfer, P., Bajracharya, D., Plachy, C., 1979. Control of Seed Germination by Abscisic Acid: I. Time Course of Action in *Sinapis alba* L 1. *Plant Physiol.* 64, 822–827. <https://doi.org/10.1104/pp.64.5.822>
- Schreiber, L., 2010. Transport barriers made of cutin, suberin and associated waxes. *Trends Plant Sci.* 15, 546–553. <https://doi.org/10.1016/j.tplants.2010.06.004>
- Schreiber, L., Franke, R., Hartmann, K., 2005. Wax and suberin development of native and wound periderm of potato (*Solanum tuberosum* L.) and its relation to peridermal transpiration. *Planta* 220, 520–530. <https://doi.org/10.1007/s00425-004-1364-9>
- Schrider, D.R., Hourmozdi, J.N., Hahn, M.W., 2011. Pervasive Multinucleotide Mutational Events in Eukaryotes. *Curr. Biol.* 21, 1051–1054. <https://doi.org/10.1016/j.cub.2011.05.013>
- Shanmugarajah, K., Linka, N., Gräfe, K., Smits, S.H.J., Weber, A.P.M., Zeier, J., Schmitt, L., 2019. ABCG1 contributes to suberin formation in *Arabidopsis thaliana* roots. *Sci. Rep.* 9, 11381. <https://doi.org/10.1038/s41598-019-47916-9>
- Shapiro, M.D., Bell, M.A., Kingsley, D.M., 2006. Parallel genetic origins of pelvic reduction in vertebrates. *PNAS* 103, 13753–13758. <https://doi.org/10.1073/PNAS.0604706103>

- Singh, B.N., Kumar, K., 1935. The influence of partial pressure of carbon dioxide on photosynthetic efficiency. *Proc. Indian Acad. Sci.* 1, 909–927. <https://doi.org/10.1007/BF03039852>
- Sözen, C., Schenk, S.T., Boudsocq, M., Chardin, C., Almeida-Trapp, M., Krapp, A., Hirt, H., Mithöfer, A., Colcombet, J., 2020. Wounding and Insect Feeding Trigger Two Independent MAPK Pathways with Distinct Regulation and Kinetics[OPEN]. *Plant Cell* 32, 1988–2003. <https://doi.org/10.1105/tpc.19.00917>
- Steinhauser, F., Eckel, O., Lauscher, F., 1958. Das Strahlungsklima, in: Steinhauser, F., Eckel, O., Lauscher, F. (Eds.), *Klimatographie von Österreich*. Springer Vienna, Vienna, pp. 13–102. [https://doi.org/10.1007/978-3-7091-5722-0\\_2](https://doi.org/10.1007/978-3-7091-5722-0_2)
- Stevens, G.C., 1992. The Elevational Gradient in Altitudinal Range: An Extension of Rapoport's Latitudinal Rule to Altitude. *Am. Nat.* 140, 893–911.
- Sun, Y.-Q., Zhao, W., Xu, C.-Q., Xu, Y., El-Kassaby, Y.A., De La Torre, A.R., Mao, J.-F., 2020. Genetic Variation Related to High Elevation Adaptation Revealed by Common Garden Experiments in *Pinus yunnanensis*. *Front. Genet.* 10. <https://doi.org/10.3389/fgene.2019.01405>
- Tranquillini, W., 1960. Das Lichtklima wichtiger Pflanzengesellschaften, in: Pirson, A. (Ed.), *Die CO<sub>2</sub>-Assimilation / The Assimilation of Carbon Dioxide: In 2 Teilen / 2 Parts*. Springer, Berlin, Heidelberg, pp. 1318–1352. [https://doi.org/10.1007/978-3-642-94798-8\\_54](https://doi.org/10.1007/978-3-642-94798-8_54)
- Vishwanath, S.J., Delude, C., Domergue, F., Rowland, O., 2015. Suberin: biosynthesis, regulation, and polymer assembly of a protective extracellular barrier. *Plant Cell Rep.* 34, 573–586. <https://doi.org/10.1007/s00299-014-1727-z>
- Vishwanath, S.J., Kosma, D.K., Pulsifer, I.P., Scandola, S., Pascal, S., Joubès, J., Dittrich-Domergue, F., Lessire, R., Rowland, O., Domergue, F., 2013. Suberin-Associated Fatty Alcohols in *Arabidopsis*: Distributions in Roots and Contributions to Seed Coat Barrier Properties. *Plant Physiol.* 163, 1118–1132. <https://doi.org/10.1104/pp.113.224410>
- Wang, Yong, Wang, M., Sun, Y., Wang, Yanting, Li, T., Chai, G., Jiang, W., Shan, L., Li, C., Xiao, E., Wang, Z., 2015. FAR5, a fatty acyl-coenzyme A reductase, is involved in primary alcohol biosynthesis of the leaf blade cuticular wax in wheat (*Triticum aestivum* L.). *J. Exp. Bot.* 66, 1165–1178. <https://doi.org/10.1093/jxb/eru457>
- Wood, T.E., Burke, J.M., Rieseberg, L.H., 2005. Parallel genotypic adaptation: When evolution repeats itself. *Genetica* 123, 157–170. <https://doi.org/10.1007/s10709-003-2738-9>
- Woolfson, K.N., Esfandiari, M., Bernards, M.A., 2022. Suberin Biosynthesis, Assembly, and Regulation. *Plants* 11, 555. <https://doi.org/10.3390/plants11040555>
- Wos, G., Arc, E., Hülber, K., Konečná, V., Knotek, A., Požárová, D., Bertel, C., Kaplenig, D., Mandáková, T., Neuner, G., Schönswetter, P., Kranner, I., Kolář, F., 2022. Parallel local adaptation to an alpine environment in *Arabidopsis arenosa*. *J. Ecol.* 1–14. <https://doi.org/10.1111/1365-2745.13961>
- Wos, G., Mořkovská, J., Bohutínská, M., Šrámková, G., Knotek, A., Lučanová, M., Španiel, S., Marhold, K., Kolář, F., 2019. Role of ploidy in colonization of alpine habitats in natural populations of *Arabidopsis arenosa*. *Ann. Bot.* 124, 255–268. <https://doi.org/10.1093/aob/mcz070>
- Wrenbeck, E.E., Azouz, L.R., Whitehead, T.A., 2017. Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat. Commun.* 8, 15695. <https://doi.org/10.1038/ncomms15695>
- Wright, K.M., Arnold, B., Xue, K., Šurinová, M., O'Connell, J., Bomblies, K., 2015. Selection on Meiosis Genes in Diploid and Tetraploid *Arabidopsis arenosa*. *Mol. Biol. Evol.* 32, 944–

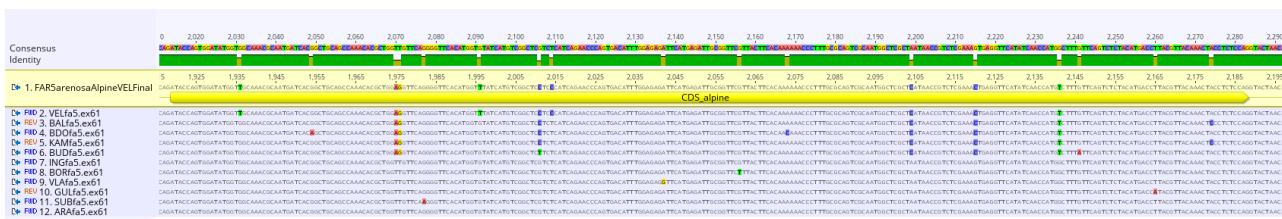
955. <https://doi.org/10.1093/molbev/msu398>

- Yang, Y., Liu, G., Guo, X., Liu, W., Xue, J., Ming, B., Xie, R., Wang, K., Hou, P., Li, S., 2022. Quantitative Relationship Between Solar Radiation and Grain Filling Parameters of Maize. *Front. Plant Sci.* 13. <https://doi.org/10.3389/fpls.2022.906060>
- Yant, L., Bomblies, K., 2017. Genomic studies of adaptive evolution in outcrossing *Arabidopsis* species. *Curr. Opin. Plant Biol.*, 36 Genome studies and molecular genetics 36, 9–14. <https://doi.org/10.1016/j.pbi.2016.11.018>
- Yeaman, S., Gerstein, A.C., Hodgins, K.A., Whitlock, M.C., 2018. Quantifying how constraints limit the diversity of viable routes to adaptation. *PLoS Genet.* 14. <https://doi.org/10.1371/JOURNAL.PGEN.1007717>
- Yeaman, S., Whiting, J., Booker, Tom, Rougeux, C., Lind, B., Singh, P., Lu, M., Huang, K., Whitlock, M., Aitken, S., Andrew, R., Borevitz, J., Bruhl, J.J., Collins, T., Fischer, M., Hodgins, K., Holliday, J., Ingvarsson, P.K., Janes, J., Khandaker, M., Koenig, D., Kreiner, J., Kremer, A., Lascoux, M., Leroy, T., Milesi, P., Murray, K., Rellstab, C., Rieseberg, L., Roux, F., Stinchcombe, J., Telford, I.R.H., Todesco, M., Wang, B., Weigel, D., Willi, Y., Wright, S., Zhou, L., 2023. Core genes driving climate adaptation in plants. <https://doi.org/10.21203/rs.3.rs-3434061/v1>
- Zheng, C., Tan, L., Sang, M., Ye, M., Wu, R., 2020. Genetic adaptation of Tibetan poplar (*Populus szechuanica* var. *tibetica*) to high altitudes on the Qinghai–Tibetan Plateau. *Ecol. Evol.* 10, 10974–10985. <https://doi.org/10.1002/ece3.6508>

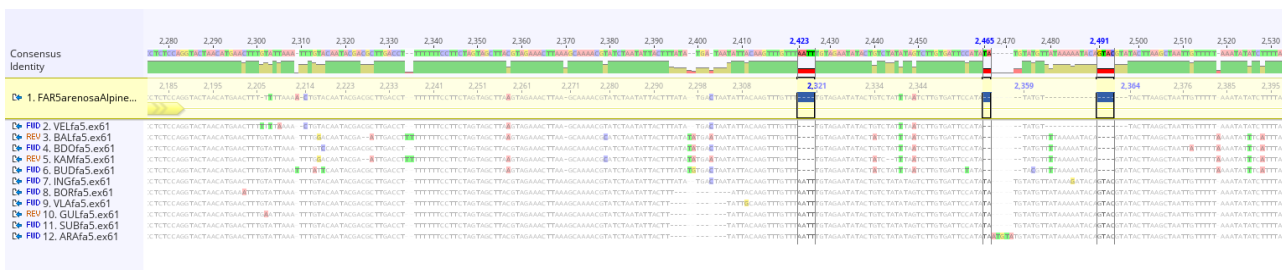
# Supplements



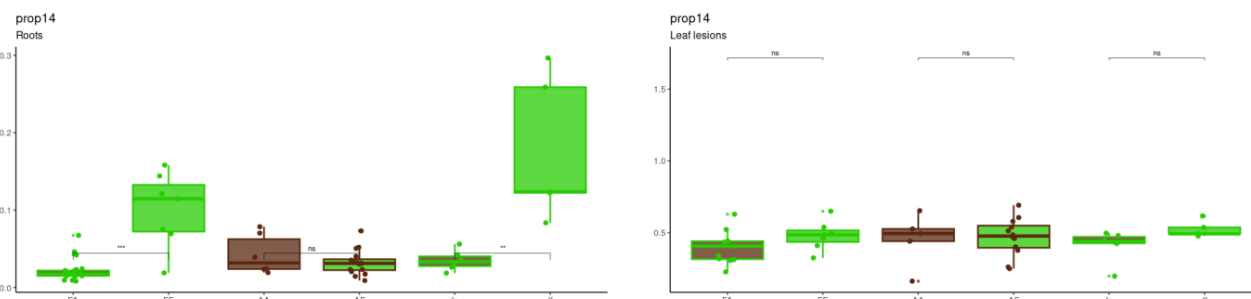
Suppl. Fig. 1: Comparison of the relative abundance of C18:0-OH between unwounded leaf (CTR) and leaves harvested after 3, 4 and 5 days after wounding (DAW). The induction of production of the fatty alcohol is strongest after 5 days.



Suppl. Fig. 2: Search for structural variants among our sequences.

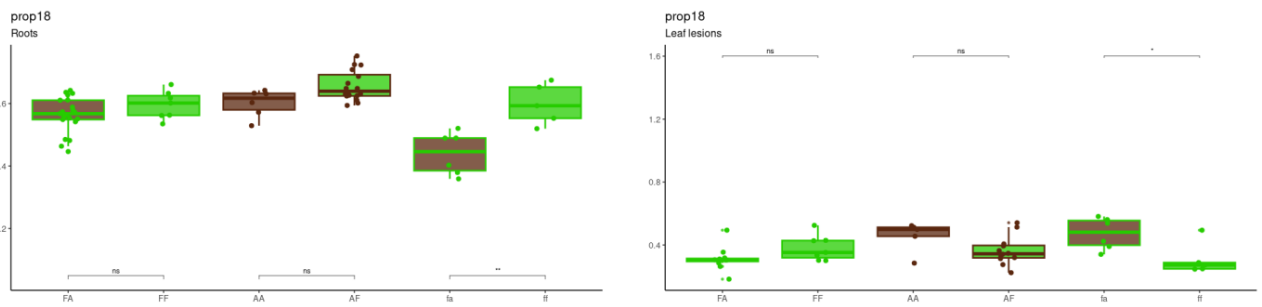


Suppl. Fig. 3: Three deletions in the intron 8.

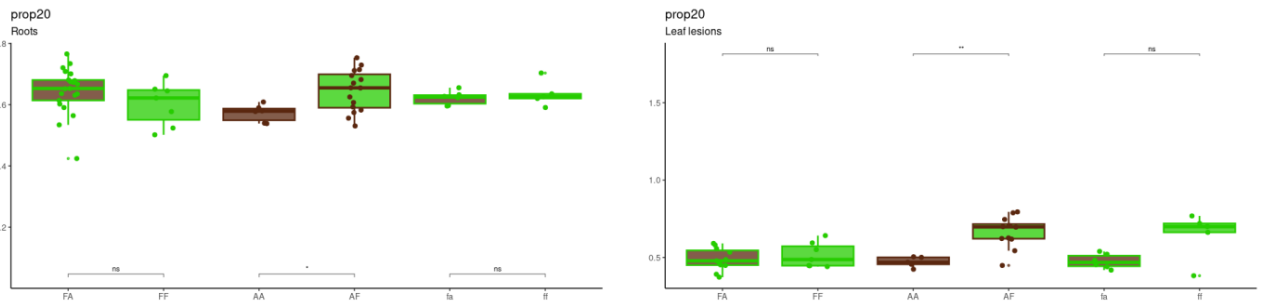


Suppl. Fig. 4: Proportion of production of tetradecanol ( $C14:0\text{-OH}/(C14:0\text{-OH}+C16:0\text{-OH})$ ) in roots (left) and wounded leaves (right).

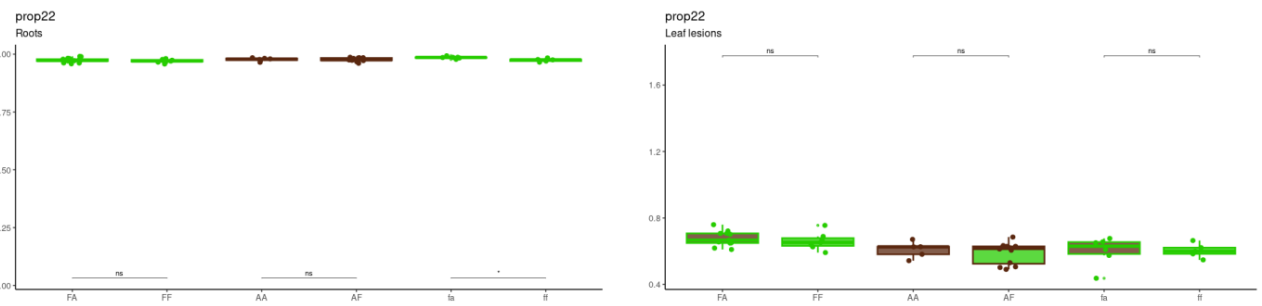




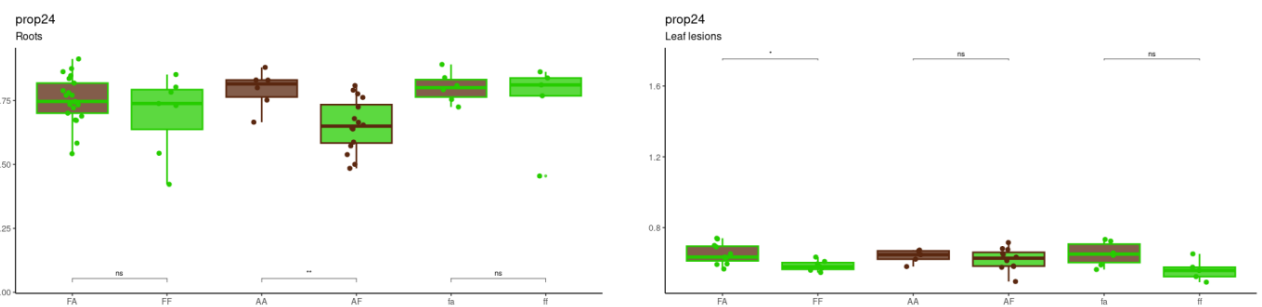
Suppl. Fig. 5: Proportion of production of octadecanol ( $C18:0-OH/(C18:0-OH+C20:0-OH)$ ) in roots (left) and wounded leaves (right).



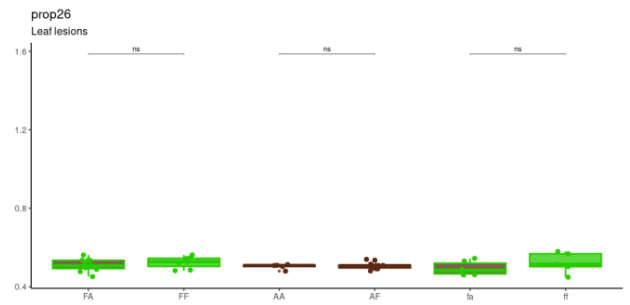
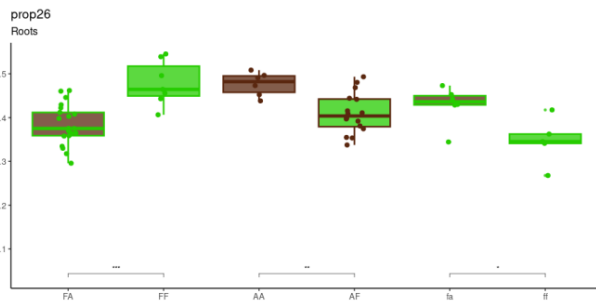
Suppl. Fig. 6: Proportion of production of eicosanol ( $C20:0-OH/(C20:0-OH+C22:0-OH)$ ) in roots (left) and wounded leaves (right).



Suppl. Fig. 7: Proportion of production of docosanol ( $C22:0-OH/(C22:0-OH+C24:0-OH)$ ) in roots (left) and wounded leaves (right).

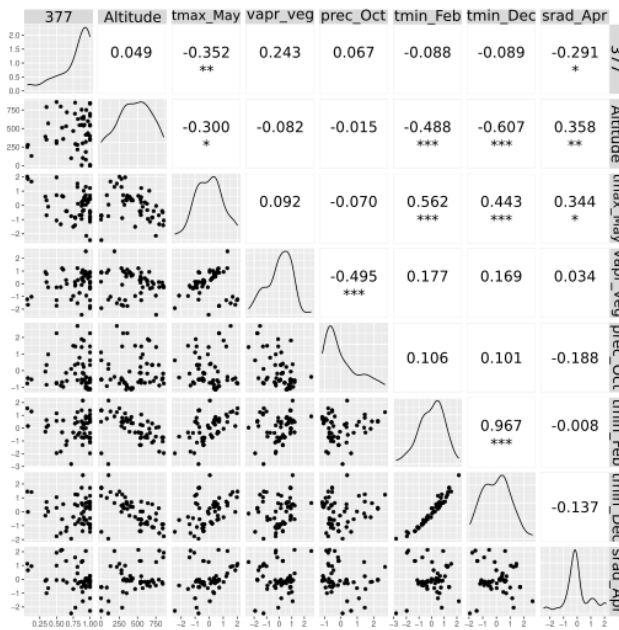


Suppl. Fig. 8: Proportion of production of tetracosanol ( $C24:0-OH/(C24:0-OH+C26:0-OH)$ ) in roots (left) and wounded leaves (right).

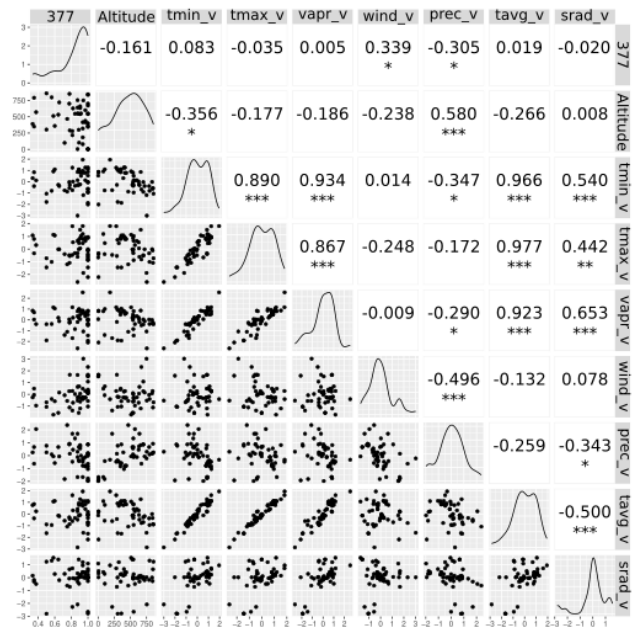


Suppl. Fig. 9: Proportion of production of hexacosanol (C26:0-OH/(C26:0-OH+C28:0-OH)) in roots (left) and wounded leaves (right).

All variables



Variables for vegetative months



Suppl. Fig. 10: Correlation of the A allele frequency (specifically position 377) across foothill populations without the Pannonian lineage with environmental variables. Correlation of the best associated variables (left) and variables for the vegetative month (right).

Suppl. Table 1: Information about the 73 populations of *A. arenosa*, Pop=code, Lin=alpine lineage, Lat=latitude, Lon=longitude, Altitude=meters above sea level, Ploidy=population ploidy, Veg=most important vegetative month, Name=local name near the pop. coordinates, State=state of locality, Nind/pop=number of individuals sampled per population, Data\_source=source publication, Bedrock=calcareous (calc), silicious (silic), serpentine (serp) or metalliferous (metal) bedrock.

Pop	Lin	Lat	Lon	Altitude	Ploidy	Veg	Name	State	Nind/pop	Data_source	Bedrock
BAB	VT	49.043514	20.180772	844	2	5	Baba	SK	12	Bohutinska et al 2021	calc
BAL	FG	45.602	24.62263889	2269	4	8	Lacul Balea	RO	8	Bohutinska et al 2021	silic
BDO		47.458	18.92477778	252	2	4	Budaors	HU	8	unpubl. Filip - mixedploidy	calc
BEL		46.16167	16.115	550	2	4	Beleograd	HR	8	Monnahan et al 2019	calc
BGS		47.62806	13.00167	570	4	5	Berchtesgaden	D	8	Monnahan et al 2019	NA
BIH		44.88181	15.89882	217	2	4	Bihac	BIH	8	Monnahan et al 2019	calc
BOR		49.6838164	15.1332558	416	4	5	Borovsko	CZ	8	Konecna et al 2021	serp
BRD		50.04967	13.89081	350	4	5	Brdatka	CZ	5	Monnahan et al 2019	silic
BUD		45.467007	25.227617	1013	2	5	Brusturet	RO	8	unpubl. Filip - mixedploidy	calc
CAR	RD	47.57594444	25.07711111	981	4	5	Carlibaba	RO	8	Bohutinska et al 2021	silic
CHO		50.59298	5.44383	103	4	5	Chokier	B	8	Monnahan et al 2019	calc
DRA	FG	45.44164	25.22394	858	4	5	Dambiovioara	RO	8	Monnahan et al 2019	calc
FOJ		43.97502	17.82446	754	2	4	Fojnica	BIH	8	Monnahan et al 2019	silic
FUG		48.6314972	15.5572367	436	4	5	Fuglau	AT	8	Konecna et al 2021	silic
GOR		44.26528	21.54271	184	2	5	Gornjak	SRB	8	Monnahan et al 2019	calc
GUL	NT	47.281679	14.927647	628	4	5	Gulsen	AT	19	Monnahan et al 2019 + Konecna et al. 2021	serp
HAR		48.85166	15.85833	400	4	5	Hardegg	AT	2	Monnahan et al 2019	NA
HLI	VT	49.17	20.03	1650	2	8	Hlinska dolina	SK	2	Novikova et al 2016	silic
HMC	VT	48.82	19.02	700	2	5	Harmanec	SK	4	Novikova et al 2016	calc
HNE		48.26694	19	280	2	4	Hotnianske Nemce	SK	7	Monnahan et al 2019	silic
HNI	VT	48.8775	20.5275	836	2	5	Hnicik	SK	4	Monnahan et al 2019	silic
HOC	NT	47.37	15.38667	545	4	5	Hochlantsch	AT	8	Monnahan et al 2019	silic
HRA	ZT	49.00716	20.286407	720	4	5	Hranovnica	SK	13	Bohutinska et al 2021	silic
HRN		62.6	18.03	5	4	5	Harnosand	S	6	Novikova et al 2016	NA
INE	RD	47.52734	24.88062	2017	4	8	Ineu	RO	7	Bohutinska et al 2021	silic
ING	NT	47.2840589	14.6815483	950	4	5	Ingeringgraben	AT	8	Konecna et al 2021	silic
KAM	ZT	49.210747	20.928184	633	4	5	Kamenica	SK	8	unpubl. Filip - mixedploidy	calc
KAS	NT	46.68833	14.87167	660	4	5	Kasparstein	AT	8	Monnahan et al 2019	calc
KLE		50.2440833	16.848417	750	4	5	Kletno	PL	9	Preite et al. 2019	metal
KOS	NT	47.74694	13.68972	467	4	5	Kosslbach	AT	7	Monnahan et al 2019	calc
KOW		50.763153	15.8439	670	4	5	Kowary	PL	8	Monnahan et al 2019	NA
KRM		50.118939	25.739772	320	2	5	Kremenets	UA	8	unpubl. Filip	calc
KZL		47.72444	18.77917	330	2	4	Kesztoic	HU	5	Monnahan et al 2019	calc
LAC	FG	45.59535	24.63458	2092	4	8	Lacul Capra	RO	8	Monnahan et al 2019	silic
MA		50.5030833	18.93816	300	4	5	Miasteczko Slaskie	PL	9	Preite et al. 2019	metal
MIE		53.92109	14.42157	5	2	5	Medzydroje	PL	11	Monnahan et al 2019	silic
OPP	NT	47.46403	14.23989	1750	4	8	Oppenberg	AT	8	Konecna et al 2021	serp
PAD		47.402365	24.546105	545	2	5	Parva	RO	9	unpubl. Filip - mixedploidy	silic
PAT	RD	47.40244	24.5459	531	4	5	Parva	RO	7	unpubl. Filip - mixedploidy	silic
PER	NT	47.35512	15.33697	540	4	5	Pernegg	AT	8	Konecna et al 2021	serp
PHD	VT	48.96228204	20.40193797	550	2	5	Prielom Hornadu	SK	8	unpubl. Filip - mixedploidy	calc
PHT	ZT	48.95281702	20.41887604	540	4	5	Prielom Hornadu	SK	8	unpubl. Filip - mixedploidy	calc
PRE		55.37821	21.03231	1	2	5	Preila	LT	8	Monnahan et al 2019	silic
RFT		48.10104	9.049581	790	4	5	Reifal	D	11	Monnahan et al 2019	calc
RZA		45.37778	22.75833	850	2	5	Retezat	RO	9	Monnahan et al 2019	silic
SCH	NT	47.27767	14.3219	2240	4	8	Schiesseck	AT	7	Monnahan et al 2019	silic
SNO	VT	49.17417	18.86167	390	2	5	Strecno	SK	6	Monnahan et al 2019	calc
SPI	ZT	48.98889	20.775	550	4	5	Spis_Drevenik	SK	13	Monnahan et al 2019	calc
STE		52.28028	16.70944	80	4	5	Stenszew	PL	8	Monnahan et al 2019	NA
STG		48.62993	15.542567	415	4	5	Steinegg	AT	8	Konecna et al 2021	serp
SUB	VT	48.96030556	20.38327778	600	2	5	Suchá Belá	SK	17	Bohutinska et al 2021	calc
SWA		48.44784	9.422422	700	4	5	Swabia_Grindel Stiege	D	10	Monnahan et al 2019	calc
SWJ		53.897702	14.298695	5	4	5	Swinoujscie	PL	3	unpubl. Levi	NA
SZI		46.80667	17.43444	130	2	4	Szigliget	HU	5	Monnahan et al 2019	silic
TBG		48.13972	8.23667	640	4	5	Triberg	D	6	Monnahan et al 2019	NA
TIS	FG	45.569999	25.608265	797	4	5	Timisu de Sus	RO	8	Bohutinska et al 2021	calc
TKO	ZT	49.20451	19.7352	1783	4	8	Tri Kopy	SK	8	Monnahan et al 2019	silic
TRD	VT	49.251596	20.206285	1380	2	7	Tristar 2x	SK	5	Monnahan et al 2019	calc
TRE	ZT	48.89417	18.04472	280	4	5	Trencin	SK	8	Monnahan et al 2019	calc
TRT	ZT	49.24932	20.20498	1700	4	8	Tristar 4x	SK	9	Monnahan et al 2019	calc
TZI	FG	46.56667	23.67417	511	4	5	Cheile Turzii	RO	10	Monnahan et al 2019	calc
VEL	VT	49.162	20.15419	1823	2	8	Velicka dolina	SK	9	Monnahan et al 2019	silic
VID		45.36392	24.63756	900	2	5	Vidraru	RO	8	Monnahan et al 2019	silic
VLA		49.7349631	15.1748464	345	4	5	Vlastejovice	CZ	8	Konecna et al 2021	silic
VOR	NT	47.49876	14.16964	1010	4	6	Vorberg	AT	8	Konecna et al 2021	silic
VYR		59.41427	30.34568	100	4	5	Vyritsa (St Petersburg)	RUS	8	unpubl. Filip	NA
WEK		48.405022	15.472906	359	4	5	Weissenkirchen	AT	8	Monnahan et al 2019	silic
WL	NT	47.3256	14.23038	2117	4	8	Wildsee	AT	8	Bohutinska et al 2021	silic
WUL		51.308	8.487028	290	4	5	Wulmeringhausen	D	8	unpubl. Lara's pilot (Ute Kramer)	metal
ZAP	ZT	49.278343	19.96706	915	4	5	Zakopane	PL	8	Monnahan et al 2019	calc
ZEP	VT	49.20652778	20.21505556	1625	2	5	Zelené pleso	SK	13	Bohutinska et al 2021	silic
ZID		46.10676	15.30565	303	2	4	Zidani Most-Gracnica	SLO	8	unpubl. Filip - mixedploidy	calc
ZIT		46.094332	15.343606	360	4	4	Zidani Most-Gracnica	SLO	8	unpubl. Filip - mixedploidy	calc

Suppl. Table 2: Outlier SNPs fro Picmin, locus=position in the genome, p=p-value, q=q-value, AL=code for *A. lyrata*, AT=code for *A. thaliana*, type=type of change, short description=description of a function

locus	p	q	AL	AT	type	short_description
scaffold_1_6591000	9.999900000099999e-06	0.0097033206883121	AL1627730	AT1618620.1	protein_coding	inn
scaffold_1_13465000	9.999900000099999e-06	0.0097033206883121	AL1644240	AT1630290.1	protein_coding	Tetrairicopeptide repeat (TPR)-like superfamily protein
scaffold_3_2962000	9.999900000099999e-06	0.0097033206883121	AL3G18260	AT3207040.1	protein_coding	NE-ARC domain-containing disease resistance protein
scaffold_3_23695000	9.999900000099999e-06	0.0097033206883121	AL3G53130	AT2181790.1	protein_coding	phytochrome B
scaffold_3_23695000	9.999900000099999e-06	0.0097033206883121	AL3G53130	AT2181790.1	protein_coding	phytochrome B
scaffold_3_23700000	9.999900000099999e-06	0.0097033206883121	AL3G53130	AT2181790.1	protein_coding	phytochrome B
scaffold_3_23701000	9.999900000099999e-06	0.0097033206883121	AL3G53130	AT2181790.1	protein_coding	phytochrome B
scaffold_3_23701500	9.999900000099999e-06	0.0097033206883121	AL3G53130	AT2181790.1	protein_coding	phytochrome B
scaffold_4_21640500	9.999900000099999e-06	0.0097033206883121	AL4G43360	AT3G10180.1	protein_coding	P-loop containing nucleoside triphosphate hydrolases superfamily protein
scaffold_5_11157500	9.999900000099999e-06	0.0097033206883121	AL5G22700	AT3G44550.1	protein_coding	fatty_acid_reductase 5
scaffold_6_11799500	9.999900000099999e-06	0.0097033206883121	AL6G37970	AT5G28000.1	protein_coding	thioglucooside glucohydrolase 1
scaffold_6_11799500	9.999900000099999e-06	0.0097033206883121	AL6G37970	AT5G28000.1	protein_coding	thioglucooside glucohydrolase 1
scaffold_6_11799500	9.999900000099999e-06	0.0097033206883121	AL6G37970	AT5G28000.1	protein_coding	thioglucooside glucohydrolase 1
scaffold_6_11799500	9.999900000099999e-06	0.0097033206883121	AL6G37980	AT5G28000.1	protein_coding	thioglucooside glucohydrolase 1
scaffold_6_11799500	9.999900000099999e-06	0.0097033206883121	AL6G37980	AT5G28000.1	protein_coding	thioglucooside glucohydrolase 1
scaffold_6_20422000	9.999900000099999e-06	0.0097033206883121	AL6G44610	AT4G10560.1	protein_coding	Oxysterol/Histidine-rich C1 domain family protein
scaffold_8_8755000	9.999900000099999e-06	0.0097033206883121	AL8G11840	AT5G46450.1	protein_coding	Disease resistance protein (TIR-NBS-LRR class) family
scaffold_8_8770000	9.999900000099999e-06	0.0097033206883121	AL8G11840	AT5G46450.1	protein_coding	Disease resistance protein (TIR-NBS-LRR class) family
scaffold_8_8920000	9.999900000099999e-06	0.0097033206883121	AL8G11850	AT2G34690.1	protein_coding	Glycolipid transfer protein (GLTP) family protein
scaffold_8_1695000	9.999900000099999e-06	0.0097033206883121	AL8G13230	AT5G45490.1	protein_coding	P-loop containing nucleoside triphosphate hydrolases superfamily protein
scaffold_8_2903000	9.999900000099999e-06	0.0097033206883121	AL8G14630	AT5G44870.1	protein_coding	Disease resistance protein (TIR-NBS-LRR class) family
scaffold_8_19194000	9.999900000099999e-06	0.0097033206883121	AL8G36160	AT5G59650.1	protein_coding	Leucine-rich repeat protein kinase family protein
scaffold_8_19194500	9.999900000099999e-06	0.0097033206883121	AL8G36160	AT5G59650.1	protein_coding	Leucine-rich repeat protein kinase family protein
scaffold_4_2790500	1.99998000002e-05	0.015644129272993	AL4G14810	AT4G36150.1	protein_coding	Disease resistance protein (TIR-NBS-LRR class) family
scaffold_5_10836000	1.99998000002e-05	0.015644129272993	AL5G32420	AT1G344290.1	protein_coding	NAC domain containing protein 80
scaffold_1_17705500	2.99997000003e-05	0.019824687958017	AL1650810	AT1G35910.1	protein_coding	Helicoid dehalogenase-like hydrolase (HAD) superfamily protein
scaffold_4_23319000	2.99997000003e-05	0.019824687958017	AL4G47730	AT2G48180.1	protein_coding	Tudor/PWWP/MBT domain-containing protein
scaffold_8_2902500	3.99996000004e-05	0.02158930051937086	AL8G14630	AT5G44870.1	protein_coding	Disease resistance protein (TIR-NBS-LRR class) family
scaffold_6_1066000	5.999940000059999e-05	0.0270551412132938	AL6G13060	AT5G03560.2	protein_coding	Tetrairicopeptide repeat (TPR)-like superfamily protein
scaffold_1_8754500	6.999930000069999e-05	0.02932200565062109	AL7G32630	AT4G28850.1	protein_coding	xyloglucan endotransglucosylase/hydrolase 26
scaffold_1_26831500	7.999920000079999e-05	0.031288258545986	AL1658810	AT1G50500.2	protein_coding	AMP-dependent synthetase and ligase family protein
scaffold_6_1620000	7.999920000079999e-05	0.031288258545986	AL6G14390	AT5G04730.1	protein_coding	Membrane trafficking VPS53 family protein
scaffold_8_1731000	8.999910000089999e-05	0.032896971608483	AL8G13270	AT5G45470.1	protein_coding	Ankyrin-repeat containing protein
scaffold_6_1065000	9.999900000099999e-05	0.0351634098337916	AL6G13060	AT5G03560.2	protein_coding	Protein of unknown function (DUF594)
scaffold_2_4154500	0.000109989890011	0.0373105560979788	AL2G17810	AT1G58400.1	protein_coding	Tetrairicopeptide repeat (TPR)-like superfamily protein
scaffold_6_1065500	0.000109989890011	0.0373105560979788	AL6G13060	AT5G03560.2	protein_coding	Disease resistance protein (CC-NBS-LRR class) family
scaffold_7_14214500	0.000119988800012	0.038977458157622	AL7G42020	AT4G14280.1	protein_coding	Tetrairicopeptide repeat (TPR)-like superfamily protein
scaffold_2_11813500	0.000139988600014	0.042251467249107	AL2G24970	AT2G43500.1	protein_coding	ARM repeat s superfamily protein
scaffold_7_21823500	0.000149988500015	0.0435546780895827	AL7G48560	AT3G58290.1	protein_coding	Plant regulator RWP-RK family protein
scaffold_7_21823500	0.000149988500015	0.0435546780895827	AL7G48560	AT3G58290.1	protein_coding	ENTH/VHS family protein
scaffold_7_8620500	0.000159988400016	0.0455947656878309	AL7G31110	AT4G22517.1	protein_coding	Class-II DAHP synthetase family protein
scaffold_7_12652000	0.000179988200018	0.0484144632237888	AL7G39550	AT5G05850.1	protein_coding	Blfunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
scaffold_7_12652000	0.000179988200018	0.0484144632237888	AL7G39550	AT5G05850.1	protein_coding	plant intracellular ras group-related LRR 1













Suppl. Table 4: Outlier SNPs from environmental GWAS, Loci=position in the genome, p.value=p-value, chromosome=chromosome number according to *A. lyrata*, gene=*A. lyrata* gene code, AT=*A. thaliana* gene code, gene\_model=type of change, gene\_name=short gene name, short\_description=description of a function, Curator\_summary=details about the function

Loci	p.value	chromosome	gene	AT	gene_model	gene_name	short_description	Curator_summary
AL7G295830.Asp900Ser.12384349	6.14E-60	2	AL2G25830	AT1G67310.1	protein_coding	CAMTA	Chromatin-binding transcription activator protein with OS-1 and Ankyrin domains	nm
AL1G36270.Asp7Val.10031288	1.08E-59	1	AL1G36270	AT1G22730.1	protein_coding	MRP2	MA3 domain-containing protein	nm
AL5G22700.Leis377Val.1157690	5.72E-58	5	AL5G22700	AT3G44550.1	protein_coding	FAR5	fatty acid reductase 5	Encodes a member of the eight-member gene family encoding alcohol-forming fatty acyl-CoA reductases (FARs) identified in <i>Arabidopsis thaliana</i> . Three of the FARs, FAR1 (A1G92500), FAR4 (A3G44540) and FAR5 (A3G44550) are shown to generate the fatty alcohols found in root, seed coat and wound-induced leaf tissue. The mRNA is cell-to-cell mobile.
AL1G36270.Glu596Ile.10031261	2.15E-57	1	AL1G36270	AT1G22730.1	protein_coding	MRP2	MA3 domain-containing protein	nm
AL5G22700.Gly263Cys.1157546	5.94E-56	5	AL5G22700	AT3G44550.1	protein_coding	FAR5	fatty acid reductase 5	Encodes a member of the eight-member gene family encoding alcohol-forming fatty acyl-CoA reductases (FARs) identified in <i>Arabidopsis thaliana</i> . Three of the FARs, FAR1 (A1G92500), FAR4 (A3G44540) and FAR5 (A3G44550) are shown to generate the fatty alcohols found in root, seed coat and wound-induced leaf tissue. The mRNA is cell-to-cell mobile.
AL7G27590.Asp433His.7115393	1.92E-55	7	AL7G27590	AT4G25500.3	protein_coding	RSP35	arginine/serine-rich splicing factor 35	Encodes an arginine/serine-rich splicing factor. The transcript is alternatively spliced and is differentially expressed in different tissues (low est, roots, stems and leaves). Barta et al (2010) have proposed a nomenclature for Serine/Arginine-Rich Protein Splicing Factors (SR proteins). Plant Cell. 2010;22:2926. RS40 binds to HYL1 and localizes to the nuclear dicing body. Along with the RS41 it appears to be involved in pre-mRNA processing and mRNA biogenesis (DOI:10.1093/nar/gkq751).
AL7G27590.Asp433His.7115393	1.92E-55	7	AL7G27590	AT4G25500.3	protein_coding	RSP35	arginine/serine-rich splicing factor 35	Encodes an arginine/serine-rich splicing factor. The transcript is alternatively spliced and is differentially expressed in different tissues (low est, roots, stems and leaves). Barta et al (2010) have proposed a nomenclature for Serine/Arginine-Rich Protein Splicing Factors (SR proteins). Plant Cell. 2010;22:2926. RS40 binds to HYL1 and co-localizes to the nuclear dicing body. Along with the RS41 it appears to be involved in pre-mRNA processing and mRNA biogenesis (DOI:10.1093/nar/gkq751).
AL7G27590.Asp433His.7115393	1.92E-55	7	AL7G27590	AT4G25500.4	protein_coding	RSP35	arginine/serine-rich splicing factor 35	Encodes an arginine/serine-rich splicing factor. The transcript is alternatively spliced and is differentially expressed in different tissues (low est, roots, stems and leaves). Barta et al (2010) have proposed a nomenclature for Serine/Arginine-Rich Protein Splicing Factors (SR proteins). Plant Cell. 2010;22:2926. RS40 binds to HYL1 and co-localizes to the nuclear dicing body. Along with the RS41 it appears to be involved in pre-mRNA processing and mRNA biogenesis (DOI:10.1093/nar/gkq751).
AL1G27640.Val1352Ile.6546873	3.98E-55	1	AL1G27640	AT1G16520.1	protein_coding	ABC40	pleiotropic drug resistance 12	ABC transporter family involved in ABA transport and resistance to lead. Localizes to plasma membrane. Upregulated by lead. Expressed in leaves, flowers, stomata and roots.
AL5G22700.Ala384Ser.1157711	8.00E-55	5	AL5G22700	AT3G44550.1	protein_coding	FAR5	fatty acid reductase 5	Encodes a member of the eight-member gene family encoding alcohol-forming fatty acyl-CoA reductases (FARs) identified in <i>Arabidopsis thaliana</i> . Three of the FARs, FAR1 (A1G92500), FAR4 (A3G44540) and FAR5 (A3G44550) are shown to generate the fatty alcohols found in root, seed coat and wound-induced leaf tissue. The mRNA is cell-to-cell mobile.
AL2G12590.Ile51Val.1293985	1.79E-54	2	AL2G12590	AT1G63010.4	protein_coding	VPT1	Major Facilitator Superfamily with SPX (SYG1/Pho81/XRR1) domain-containing protein	nm
AL1G36270.Glu60Arg.10031475	4.29E-54	1	AL1G36270	AT1G22730.1	protein_coding	MRP2	MA3 domain-containing protein	nm
AL4G31620.Asp99Gly.16861419	4.08E-54	4	AL4G31620	AT2G35585.1	protein_coding	CFTR	cystic fibrosis transmembrane conductance regulator	Encodes an SPX domain protein that transports Pi into the vacuole and is essential for phosphate homeostasis.
AL2G12590.Leu654Met.1293984	7.04E-54	2	AL2G12590	AT1G63010.4	protein_coding	VPT1	Major Facilitator Superfamily with SPX (SYG1/Pho81/XRR1) domain-containing protein	nm
AL1G36270.Val88Ile.10031705	1.31E-53	1	AL1G36270	AT1G22730.1	protein_coding	MRP2	MA3 domain-containing protein	nm
AL4G31620.Asp483Tyr.16861400	6.89E-53	4	AL4G31620	AT2G35585.1	protein_coding	CFTR	cystic fibrosis transmembrane conductance regulator	Encodes an SPX domain protein that transports Pi into the vacuole and is essential for phosphate homeostasis.
AL7G35570.Glu14Lys.10656658	1.14E-52	7	AL7G35570	AT4G18680.1	protein_coding	DOG1	delay of germination protein	nm
AL4G31620.Asp483Tyr.16861401	1.04E-52	4	AL4G31620	AT2G35585.1	protein_coding	CFTR	cystic fibrosis transmembrane conductance regulator	nm
AL7G22550.Ala688Ser.5158684	1.57E-50	7	AL7G22550	AT4G29750.1	protein_coding	CRM	CRS1/YhbY (CRM) domain-containing protein	nm















**Preparation of the botanic sample:**

1. Take approximately 1 cm<sup>2</sup> of fresh leave.
2. Place the leaves in a tea filter bag.
3. Allow it to dry over 5 days in a sealed plastic bag/box with silica gel balls.

**DNA Extraction:**

4. Beat the tubes on the (Qiagen) TissueLyser for 2-4 minutes with 50 oscillations/s until all samples are powdered.
5. Prepare the extraction buffer (EP) for 24 samples: Mix 41.6ml Extraction buffer (EP) with 41.6µl Mercaptoethanol.
6. Add a pinch of Polyvinylpyrrolidon<sup>1</sup> K30 (PVP), 1300µl of the mix (EP and Mercaptoethanol) and 5µl RNase to each tube. Mix thoroughly by hand.
7. Let the samples stand for 20 minutes at room temperature.
8. (Meanwhile, prepare new 1.5ml Eppendorf tubes for the DNA washing steps and label them.)
9. Centrifuge for 5 minutes at 7000rpm.
10. Discard the supernatant.
11. Add 300µl EP and 300µl lysis buffer (LP) to each tube. Shake well to prevent pellets from settling.
12. Incubate for 15 minutes at 65°C in the thermoblock with mixing at 300rpm.
13. Add 600µl of Chloroform (Chloroform: Isoamylalcohol 1:24) and manually mix for 1 minute.
14. Centrifuge for 10 minutes at 9,000 rpm.
15. Transfer 550µl of the upper aqueous phase to the new 1.5ml tubes (from step 8).
16. Add 370µl frozen Isopropanol to each tube.
17. Gently mix and place the tubes in the freezer (-20°C) for 30 minutes.

**DNA-Washing:**

18. Centrifuge for 15 minutes at 13,000rpm and 4°C.
19. Carefully drain the supernatant, ensuring the pellet is not flushed out.
20. Add 700µl 80% ethanol at room temperature and mix on the Multirotator (Grant-bio) for 3 minutes.
21. Centrifuge for 2 minutes at 13,000 rpm.
22. Carefully pour out the supernatant; ensure the pellets are free of ethanol. Dry the samples in opened tubes on a 60°C heating block for 6-10 minutes.
23. Add 40-100µl nuclease-free H<sub>2</sub>O and dissolve DNA in a heating block at 60°C for 10 minutes or overnight in the refrigerator at +4°C.
24. Before DNA concentration measurements, vortex the samples well.

---

<sup>1</sup> Polyvinylpyrrolidon removes secondary metabolites

## How to prepare the Extraction Buffer (EP) and the Lysis Buffer (LP)

### Extraction Buffer (EP)

			<i>For 200ml</i>	<i>For 250ml</i>
<i>0.1M</i>	<b>TRIS-HCL, pH 7,5</b>	<i>157.64 g/mol</i>	<i>3.153 g</i>	<i>3.866 g</i>
<i>0.005M</i>	<b>EDTA pH 8</b>	<i>292.25 g/mol</i>	<i>0.292 g</i>	<i>0.365 g</i>
<i>0.35M</i>	<b>Sorbitol</b>	<i>182.18 g/mol</i>	<i>12.753 g</i>	<i>15.94 g</i>
<i>10mM</i>	<b>2-Merkaptoethanol (0.1%)</b>		<i>0.2 ml</i>	<i>0.25 ml</i>

#### Calculation for 250ml Extraction Buffer:

**Formula:**  $m = c \cdot V \cdot M$

**TRIS-HCL:**  $0.1 \text{ mol/L} \cdot 0.250 \text{ L} \cdot 157.64 \text{ g/mol} = \underline{3.9 \text{ g}}$

**EDTA:**  $0.005 \text{ mol/L} \cdot 0.250 \text{ L} \cdot 292.25 \text{ g/mol} = \underline{0.37 \text{ g}}$

**Sorbitol:**  $0.35 \text{ mol/l} \cdot 0.250 \text{ L} \cdot 182.18 = \underline{15.94 \text{ g}}$

**Information:** Make sure that the EP is weighed without the addition of mercaptoethanol. To adjust the water to pH 8, add NaOH. Heat Tris-HCl and EDTA slightly to facilitate dissolution. After cooling, add the mercaptoethanol.

### Lysis Buffer (LP)

			<i>For 100ml</i>	<i>For 200ml</i>	<i>For 250ml</i>
<i>0.2M</i>	<b>TRIS-HCL, pH 7.5</b>	<i>157.64 g/mol</i>	<i>3.15g</i>	<i>6.306g</i>	<i>7.88g</i>
<i>0.05M</i>	<b>EDTA pH8</b>	<i>292.25 g/mol</i>	<i>1.46g</i>	<i>2.923g</i>	<i>3.65g</i>
<i>2M</i>	<b>NaCl</b>	<i>58.44 g/mol</i>	<i>11.69g</i>	<i>23.376g</i>	<i>29.22g</i>
<i>2%</i>	<b>CTAB</b>	<i>378,5 g/mol</i>	<i>2g</i>	<i>4g</i>	<i>5g</i>

#### Calculation for 250ml Lysis Buffer:

**Formula:**  $m = c \cdot V \cdot M$

**TRIS-HCL:**  $0.2 \text{ mol/L} \cdot 0.250 \text{ L} \cdot 157.64 \text{ g/mol} = \underline{7.882 \text{ g}}$

**EDTA:**  $0.05 \text{ mol/L} \cdot 0.250 \text{ L} \cdot 292.25 \text{ g/mol} = \underline{3.65 \text{ g}}$

**NaCl:**  $2 \text{ mol/l} \cdot 0.250 \text{ L} \cdot 58.44 = \underline{29.22 \text{ g}}$

**CTAB:**  $250 \text{ g/100} \cdot 2 = \underline{5 \text{ g}}$

**Information:** Achieve a pH of 8 by adding NaOH to the water. Gently warm Tris-HCl and EDTA, not exceeding 65°C, to facilitate dissolution. It is advisable to add them sequentially for better dissolution, noting that EDTA dissolves more effectively when the pH is already close to 8.