# Identification of Arabis alpina genomic regions associated with climatic variables along an elevation gradient through whole genome scan

Stéphane Lobréaux, Christian Miquel

# Identification of *Arabis alpina* genomic regions associated with climatic variables along an elevation gradient through whole genome scan

Authors : Stéphane Lobréaux[1*], Christian Miquel[1]

[1] Laboratoire d'Ecologie Alpine, CNRS, Université Grenoble-Alpes, Grenoble, France

[*] corresponding author : stephane.lobreaux@univ-grenoble-alpes.fr

Keywords : SNP, local adaptation, plant, membrane-bound transcription factor, genome scan

Data accessibility : Sequence data from this article have been deposited with the European Nucleotide Archive under the project number PRJEB32228.

Abstract

We performed a pooled whole-genome sequencing on samples of the alpine plant *Arabis alpina*, harvested in ten populations along an elevation gradient in the French Alps. A large dataset of genetic variations was produced as single nucleotide polymorphisms (SNPs). A combined genome scan approach enabled detecting genomic regions associated with a synthetic environmental variable characterizing the climate at each sampling location. Positive loci detected by two methods were retained and belong to 19 regions in the *Arabis alpina* genome. The most significant region harbors an ortholog of the AtNAC062 gene, encoding a membrane-bound transcription factor described as linking the cold response and pathogen resistance that may confer protection to plants under extended snow coverage at high elevations. Other genes involved in the stress response or in flowering regulation were also detected. Altogether, our results indicated that *Arabis alpina* represent a suitable model for studying genomic adaptation in alpine perennial plants.

## 1. Introduction

Local adaptation is a process by which a population adapts to its local biotic and/or abiotic environmental conditions and becomes well adapted to this particular environment [1]. Local adaptation is an important field of research that has been stimulated by the recent development of new methods and the interest in ongoing climatic change [2, 3]. At the genome level, genetic variations related to local adaptation tend to be maintained into the population, and the corresponding genomic regions are therefore subjected to selective pressure [4]. At these loci, a change in allelic frequencies can be detected in response to environmental constraints, and different methods based on statistical tools have been developed to look for such genomic regions [5, 6]. In searching for genome-environment associations, it is important to be aware of confounding parameters, such as population structure and demography, which may lead to detecting false positives [4]. Evolution in sequencing technologies has enabled to acquire genomic data and genetic variations at a larger scale more easily and at a reduced cost [7]. Some techniques targeting genomic regions by reduced representations such as GBS [8] or RADseq [9] allow further reducing genotyping cost. However, genome resequencing remains, when affordable, a valuable strategy for obtaining SNP data at high density and along the entire genome [10] to get a more accurate and exhaustive view of genetic variations associated with environmental data. An alternative strategy for acquiring allelic frequency estimates from populations is pooled resequencing of individuals [11-13]. This strategy makes it possible to significantly reduce sequencing costs while SNP density and genome coverage are preserved. Different studies have validated and successfully implemented this strategy in the past years to investigate local adaptation in plants, for example in *Zea mays* [14] or *Populus* trees [15].

Mountain areas are suitable environments to look for local adaptation since contrasting climatic conditions occur within a rather short distance from each other [16]. It is well described

that along elevation gradients, living organisms face harsher climatic conditions with an increase of elevation. Studying populations of a species along such gradients may help to identify genomic regions involved in the local adaptation of a particular species. Due to their immobility, plants are particularly exposed to local climate conditions, and their adaptation is therefore vital. Alpine plants face exposure to stress factors such as UV, cold, high light, wind, and extended snow coverage [17]. Among alpine plants, some of them, like *Arabis alpina*, are able to grow in a wide range of elevations. This plant of the Brassicaceae family is found in the French Alps at elevations varying from 500 to 3,000 m. Such a large range of climatic conditions makes it possible to study local adaptation by comparing genomic data from populations sampled along such a climatic gradient. *Arabis alpina* has a relatively small genome of 337 Mb for which data are available [18, 19], and the sequenced genome has recently been improved using a combination of optical mapping and chromosome conformation capture [20]. This plant has emerged as a model, closely related to *Arabidopsis thaliana*, allowing to study the regulation of flowering in a perennial plant [21, 22]. But its capacity of growing in a large range of elevations makes it also very interesting for studying local adaptation. Previous studies have been conducted to detect genomic regions in *Arabis alpina* associated with local environments [23, 24]. Poncet *et al.* [23] searched for ecologically relevant loci associated with climatic variables based on samples harvested in the French and Swiss Alps. Both studies were based on amplified fragment length polymorphisms (AFLP), and with this method, a limited number of loci were investigated in the genome by a few hundred of AFLP markers.

In this study, we have investigated *Arabis alpina* local adaptation at the whole genome level by resequencing pools of individuals sampled in populations along an altitude gradient in the French Alps.

## 1. Materials and methods

### *1.1 Sampling*

The ten sampling sites are located in the French Alps, covering a range of elevations from 860 to 2,781 m (see Table1 for sampling site data). Plant material was harvested from all sites during a 15-day period in the summer. Fresh leaf samples were collected and dehydrated in silica gel, in which they were stored until DNA extraction.

### *1.2 DNA extraction and sequencing*

Genomic DNA was extracted from ten samples from each sampling location. For each sample, twenty milligrams of dried leaves were grounded into a fine powder, and DNA extraction was performed using the NucleoSpin Plant II kit (Macherey-Nagel, Duren, Germany) according to the manufacturer's instructions. Extracted DNA was quantified using the Qubit dsDNA BR assay kit on a Qubit fluorometer (Invitrogen, USA). Agarose gel electrophoresis enabled checking DNA quality and verifying that the quantification was accurate. DNA pools were prepared by mixing equal amounts of genomic DNA from the 10 samples of a location.

DNA sequencing using Illumina technology was performed by Fasteris (Geneva, Switzerland). To prepare libraries from each DNA pool, DNA was broken, and bead size selection enabled recovering inserts of 350bp. Libraries were prepared using the Illumina Genomic Nano kit. Sequencing was performed using multiplexing on a HiSeq2500 sequencer producing reads of 150bp as paired-end data. Libraries were loaded on 3 flow cell lanes to get an average genome coverage of 28x.

### *2.3 SNP mapping*

Raw reads were filtered to remove sequences with an average quality below 25, and 5' and 3' trimming was performed to delete bases with quality below 20. Any read presenting a length

lower than 50 bases after quality trimming was removed. Prior to read mapping on *Arabis alpina* chromosomes, a filtering step to get rid of abundant chloroplast reads was performed. Paired-end reads were mapped using BWA 0.7.14 [25] on *Arabis alpina* chloroplast genome (accession number HF934132, [26]). Then, IDs from all matching reads were collected, and the corresponding paired-end reads were removed from the dataset.

The resulting filtered dataset was mapped onto the *Arabis alpina* chromosomes. The last published version 5 of this genome was used in this study [20]. Mapping was performed using BWA 0.7.14, while samtools 1.3.1 and bcftools 1.3.1 enabled SNP calling and filtering [27]. SNPs located within a distance of 10 bases around an indel were filtered from the SNP dataset. We removed positions where more than 2 alleles were detected, where the coverage was lower than 10, or where the mapping quality was below 30. We retained only SNPs for which data were available in all 10 studied populations. When SNP variability was detected in only one location, the corresponding SNP was discarded from the final dataset. A SNP density filtering step was performed in which SNPs were removed when 3 SNPs or more were found within a 10 base region.

### *2.4 Genetic distance and phylogenetic tree*

One thousand SNPs were selected randomly in the mapping dataset from each of the 8 *Arabis alpina* chromosomes. Allelic frequencies of these 8,000 SNPs in the studied populations were used to generate a phylogenetic tree by the neighbor-joining method using POPTREEW [28]. The corresponding data in Newick format enabled plotting the tree with the ape R package [29].

### *2.5 Detection of loci associated with environmental variables*

#### *2.5.1 Environmental variables*

Climatic data were downloaded from the WorldClim database (http://www.worldclim.org) at a spatial resolution of 30 arcsec. Precipitations and minimum and maximum temperature data were extracted using the R package raster according to the GPS coordinates of the sampling locations (Table 1). From these monthly data, annual precipitations and the annual range of temperatures were calculated for each sampling site.

Snow cover data were downloaded from the National Snow and Ice Data Center (http://nsdic.org). Such data are based on the difference in reflectance of snow-covered land, which is high in the visible band but low in shortwave infrared. MODIS/Aqua Snow Cover 8-Day L3 Global 500m Grid Version 6 files for the 2010-2016 period were used. This dataset reports the maximum snow cover extent during an eight-day period in 1,200 km x 1,200 km tiles. According to the GPS coordinates of the sampling sites, the appropriate tiles were defined using the MODLAND tile calculator (https://landweb.modaps.eosdis.nasa.gov/cgi-bin/developer/tilemap.cgi). An average frequency of snow cover was calculated for each week throughout the period 2010-2016. The number of weeks without snow was then defined, corresponding to the sum of weeks where the snow cover frequency was below 0.5.

Principal component analysis (PCA) of environmental variables was performed with R package ade4 [30].

*2.5.2 Generalized linear mixed models (GLMM)*

GLMMs were used to detect SNPs correlated with environmental variables as previously described [31]. A GLMM model with logit link and a binomial error distribution was estimated between each SNP and the environmental variable. The R lme4 package was used to calculate the models [32]. Using the allelic frequencies of the alleles at a defined

genomic position, the number of occurrences of each allele was calculated for a total count of 30, and these occurrence data were used as input for model determination. This strategy enabled us to perform such calculations based on an equal total count of alleles at each location at a genomic position. To evaluate model performances, the likelihood ratio (LR) was calculated using the ANOVA function from the R car package with the model and null model as input. A Bonferroni correction [33] was applied to the LR p-values as previously described [31]. Wald1 and Wald2 coefficients were also used to evaluate model performances and were recovered from the GLMM model data produced [34].

*2.5.3 Latent factor mixed models (LFMM)*

The LFMM 1.5 program enables testing for the association of genomic loci with environmental variables while taking into account the population structure through unobserved latent factors ([35], http://membres-timc.imag.fr/Olivier.Francois/lfmm/index.htm). The number of latent factors is fixed using the K parameter. To define the optimal value, we tested values from 2 to 6 with a reduced dataset containing a total of 120,000 SNPs, 15000 per chromosome. Genotyping data were provided as allelic frequencies per population. LFMM implement an MCMC algorithm, the number of iterations in the Gibbs sampling algorithm was set to 30,000, and the number of burning iterations to 15,000. For the complete dataset, we performed 6 independent runs, and the Z-scores were combined using the Fisher-Stouffer method. The p-values were adjusted as described by LFMM manual (http://membres-timc.imag.fr/Olivier.Francois/lfmm/index.htm).

## 2. Results and discussion

### *2.1 Genomic sequencing and population structure analysis*

In order to acquire genomic variation data from populations of *Arabis alpina* growing in contrasting environments, we sampled leaves from individual plants along an elevation gradient in a range of 860 to 2,781 m (Table 1).

Extracted DNA was submitted to pool sequencing to get the allelic frequencies at a large number of loci across the genome, each pool containing an equivalent amount of 10 independent individuals (Supplementary data, Table S1 for sequencing statistics). A sampled population contains between 20 and 30 individual plants depending on the sampling sites, with only plants that are at least 1 meter apart being sampled since *Arabis alpina* tends to propagate at short distances through sucker production. Ten sampled plants represent therefore a significant part of each studied population, suitable for allelic frequency estimation. SNP mapping using the 8 *Arabis alpina* chromosomes that have been sequenced as a reference [20] produced a dataset of 2,575,324 SNPs for which data are available in all sampled populations. This number of SNPs corresponded to a density of one SNP at every 130 bp and well covered the whole sequenced genome of *Arabis alpina*. A phylogenetic tree was built from a subset of 8,000 SNPs to detect the genetic relationship between the sampled populations (Fig. 1). A clear genetic structure was detected, and 3 groups of populations were identified. Two groups correspond to a set of samples from the Chartreuse and Vercors area, that are pre-alpine massifs of the French Alps. A third group corresponds to samples collected in the French Alps. Interestingly, samples in the lower and higher range of elevation of the altitude gradient were collected in all 3 areas, which may help to prevent the detection of false positives when looking for loci associated with climatic variables by limiting population effects.

*2.2 Combined genome scan*

To detect associations between loci as SNPs and the climatic environment, we prepared a synthetic variable summarizing the climatic conditions in which these plant populations grow. For this purpose, a PCA was performed using 3 available variables (annual range of temperatures, annual precipitations, and snow coverage). The first axis of the PCA explained 97% of the variance of the data and was chosen as the climatic variable for the *Arabis alpina* sampling sites. Two independent analyses were then performed with statistical tools that enable us to detect associations between loci and the environmental variables.

We first used GLMM, which we have previously been shown to be effective in detecting associations between minimum temperature and SNPs in *Arabidopsis* [31]. Indeed, the analysis of 80 whole-genome data from 80 *Arabidopsis* ecotypes allowed identifying several significant loci, most of them documented as being involved in stress or cold response in *Arabidopsis thaliana*. We generated a model for each *Arabis alpina* SNP of the dataset, using the 3 groups identified in the phylogenetic tree as population structure data to take into account that parameter and reduce the detection of false positives. The presence of low and high elevation sites in each group may also contribute to reducing false positives through the population effect.

SNPs that were significantly associated with the synthetic climatic variable used in this study were selected using LR and a Wald test according to the procedure previously described [31]. At a threshold of 0.01 and after a Bonferroni correction, 533,699 SNPs were considered as positive. The lowest LR p-value in this analysis was $1.497\ 10^{-83}$, well below the threshold used.

In a second approach, we used the LFMM program to test for the association of SNP allelic frequency with the climatic variable. Six independent analyses were performed and combined as suggested by the program developers [35]. In this case, the software itself analyzed the population structure and included this parameter in the analysis. The K latent factor was fixed at 3 as the optimal value, which was in agreement with the 3 groups of populations identified in the phylogenetic analysis (Fig. 1). A 0.01 threshold led to the identification of 141,392 positive SNPs, with a lowest adjusted p-value of $2.7\ 10^{-9}$.

In order to extract from these two analyses the most significantly associated SNPs, and considering the large number of putative positive SNPs detected at the thresholds used , we selected the 500 SNPs corresponding to the lowest p-values from each method. We then combined the two sets of results. The SNPs detected by both methods were selected. Subsequently, those SNPs were retained that correspond to genomic regions for which at least two positive SNPs were detected within a 5kb window. Such a sliding-window analysis was successfully used to detect SNPs associated with minimum temperature in *Arabidopsis thaliana* [31].

*2.3 Analysis of detected genomic regions*

*2.4 A major genomic region located on chromosome 5*

Among the 19 regions detected (Supplementary data, Table S2), the most significant was located on *Arabis alpina* chromosome 5. All the 25 first LFMM lowest p-values are located within a 25kb genomic region of this chromosome, from position 11,521,406 to 11,546,695. Furthermore, this part of chromosome 5 was also the most significant one with the GLMM approach, with 13 hits within the 21st lowest p-values generated by this method. Therefore, a

strong positive signal was detected with both methods. This genomic region harbors 4 annotated genes (Fig. 2).

Aa_G315180, described as the ortholog of the *Arabidopsis thaliana* At3g49490 gene, encodes an uncharacterized protein.

Aa_G315190 and Aa_G315200 are annotated as pyruvate decarboxylase genes (PDC). In *Arabidopsis*, a family of 3 *PDC* genes has been described [36]. They encode proteins of 592 to 607 amino acids long, which is also the average size for this enzyme in other plants. The putative *Arabis PDC1* and *2* genes, however, encode much shorter proteins of respectively 111 and 306 aa. When compared to the structure of the *Arabidopsis PDC* genes, the *PDC1* encoding region is interrupted by a stop codon in the 4th exon. For *PDC2*, the whole gene structure is detected in this part of the *Arabis* chromosome 5; however, a stop codon leads to a truncated protein. We have previously assembled *Arabis alpina* genomic sequences, and the contig 7180002427937 of this assembly corresponds to this part of chromosome 5. These data confirm the sequence documented in the *Arabis alpina* genome [18], and *Arabis PDC1* and *2* are therefore very likely to be pseudogenes.

Aa_G315210 is annotated as the ortholog of the *Arabidopsis thaliana* gene At3g49530, which encodes a transcription factor named NAC062. The structure of the two genes is similar, with 6 exons and the corresponding proteins of 468 and 457 amino acid sharing 77% identity at the amino acid level. Among the SNPs detected in the Aa_G315210 gene, most of them correspond to synonymous mutations or are located in non-coding regions, and one of them is a non-synonymous mutation. At position 182 of the protein, the two alleles lead to a shift from a proline to an arginine (Fig. 3). This SNP represents therefore a non-conservative substitution in the *Arabis* NAC062 protein and its allelic profile along the elevation gradient was detected as highly correlated to the synthetic variable used. Alignment of *Arabidopsis* and *Arabis*

NAC062 protein sequences revealed that position 182 of the *Arabis* protein corresponds to an amino acid insertion compared to the counterpart in *Arabidopsis*. This is also true when *AaNAC062* is compared with its ortholog from *Capsella rubra* or *Arabidopsis lyrata*, other plants from the Brassicaceae family. It was therefore not possible to investigate whether similar mutations at this position occur in the different *Arabidopsis* ecotypes for which genomic data are available. Fig. 4 shows an example of an allelic profile that we detected as being highly correlated with the environmental variable within the Aa_G315210 gene. A clear shift in allelic frequency is detected along the gradient, with distinct alleles present at the bottom and the top part of the climatic gradient.

The Aa_G315210 gene encodes a protein of the MTF family (membrane-bound transcription factors) annotated as the ortholog of the NAC062/NTL6 protein of *Arabidopsis thaliana*. MTFs are membrane-anchored proteins that respond to stimulation by migrating to the nucleus, where they regulate the expression of their target genes [37] . Such transcription factors are involved in the stress response of different organisms [38], allowing a rapid transcriptional regulation under changing environmental conditions [39, 40]. In *Arabidopsis*, MTFs have been shown to be involved in the response to different stress factors such as high salt and cold [41-43] and developmental signals [44, 45].

The AtNAC062 protein is activated in response to cold. A primary event upon exposure to low temperatures is a change in membrane fluidity that induces the proteolytic processing of NAC062 and its release from the plasma membrane [46, 47]. NAC062 activates the expression of a set of PR proteins (pathogenesis-related proteins), linking the cold stress and the pathogen responses [41].

In winter, snow acts as a protection against freezing temperatures but maintains also a high humidity favorable for psychrophilic pathogens to attack dormant plants [48, 49]. This has been shown for overwintering plants that need to keep vegetative tissues under the snow. Long-term snow cover leads to the depletion of plant carbohydrate reserves, so plants become more susceptible to pathogen attack [50]. In these overwintering species, PR proteins protect against pathogen attacks and also provide freezing tolerance [51]. The perennial plant *Arabis alpina* also keeps vegetative tissue under the snow cover to allow for rapid flowering and seed production after the snow has melted in the spring at high elevation [52]. Along the studied climatic gradient, periods of growth without snow varies in a large range from 9 to 42 weeks at high elevations and may lead to specific adaptation to face this constraint.

*2.5 Other genomic regions associated to the synthetic climatic variable*

Among the other regions detected as being significantly associated with the synthetic climatic variable used in this study, some contain genes linked to the stress response (Table 2). The Aa_G89730 gene encodes a protein of the pfkb-like carbohydrate kinase family. Such proteins have been shown to be involved in the drought response in *Phaseolus vulgaris* [53]. On chromosome 8, an acyl-activating enzyme gene is located within the detected region. This family of enzymes perform carboxylic acid activation and is required in different metabolic pathways [54]. In *Arabidopsis*, Rizhsky *et al.* [55] have shown, through transcriptome analysis, that the expression of genes of this family is modulated by a combination of abiotic stress factors. They exposed plants simultaneously to drought and heat. Such situations of exposure to multiple stress factors are very likely to occur in mountain areas. Gene Aa_G18650 encodes a heat shock protein.

An ortholog of the *Arabidopsis thaliana AtABI3* gene was detected on chromosome 3 (abscisic acid-insensitive 3). ABI3 protein has been initially characterized as a

transcription factor involved in plant seed maturation. But recently some other functions have been reported, such as a role in plant development and flowering time [56]. Interestingly, another significant region contains an ortholog of the *AtPRMT10* gene on *Arabis alpina* chromosome 8. It encodes a protein called arginine methyltransferase, which catalyzes the asymmetric demethylation of arginine 3 of histone 4. This enzyme is involved in the regulation of flowering time [57]. The *Arabidopsis* PRMT10 mutant has a late flowering time phenotype. In *Arabis alpina*, flowering is linked to cold since a prolonged exposure to low temperatures is required to initiate the formation of flower buds [58].

Members of other gene families were also detected. An aluminum-activated malate transporter gene, encoding a protein of a family that has first been characterized in wheat as being involved in aluminum resistance, but these proteins perform many other functions as well [59]. A pectate lyase encoding enzymes important in various plant developmental processes was also detected [60]. Finally, some genes for which no functional data are available so far, such as hypothetical proteins or the ortholog of an *Arabidopsis* kinase-like protein (TMKL1) for which no cellular function has been described yet [61], were also found. Some regions do not contain annotated genes according to the actual knowledge of the *Arabis alpina* genome, but we cannot rule out that they may play a role in *Arabis alpina* adaptation.

## 3. Discussion

Genomic studies have enabled investigating local adaptation at the molecular level by looking for genome-environment associations in plant species [3]. Major contributions have been made using the model plant *Arabidopsis thaliana*, for which abundant resources are available [62-65]. In addition to sequenced genomes, functional data and characterized ecotypes from different area allow detailed investigations. The recent evolution in sequencing

techniques enables to generate high-level genomic resources in non-model organisms, opening the possibility of studying local adaptation in other species using high-density data of genetic variations. Although *Arabidopsis thaliana* has been useful in providing some knowledge of local adaptation in plants, other models may enable addressing specific questions that cannot be investigated in *Arabidopsis*. *Arabis alpina* has become a complementary species to the annual plant *Arabidopsis* to study flowering time regulation in a perennial plant [22]. Since the two plants are closely related in the Brassicaceae family, *Arabis alpina* studies benefit from functional gene data available from *Arabidopsis thaliana* resources. In this work, we investigated this species as a suitable model for studying local adaptation of alpine plants. Temperature is a well-known key factor for plant growth, and association studies with climatic variables in *Arabidopsis* have confirmed that it is an essential parameter [62, 65]. The increasing duration of the growing season, which in part is correlated with local temperature, is also an important factor for *Arabidopsis* [64]. The synthetic variable that we used in our study integrates these two factors with the annual precipitations.

If these variables are common drivers in local adaptation of plants, the major loci that we identified appear to be linked to the life cycle of a perennial alpine plant growing in contrasting environments along an elevation gradient. For positive SNPs associated with the synthetic environmental variable, we detected a shift in alleles along the elevation gradient. These results suggest that environmental pressures indeed shape genetic diversity along the gradient, leading to local adaptations. Alpine plants are mainly perennials and adapt to the specific climatic conditions where they grow by initiating flowering in response to cold through vernalization [58, 66]. In this way, flowering is restricted to spring at the beginning of the short growing season so that seeds and next plantlets are produced with optimal chances of survival. Apart from genes linked to stress responses, some major candidate genes that we identified are potentially involved in adaptations to variations in snow cover and therefore growing period;

these variations may be involved in plant protection under long-term snow cover. Flowering time is also crucial and is linked to low temperatures in *Arabis* [58], and candidate loci associated with flowering regulation were detected. Some candidate genes are therefore strongly related to key drivers of plant adaptation in mountain areas. Selective pressure associated with local climate acting through temperature and growing season varies along the studied gradient, with an increase in selective pressure at high elevation.

*Arabis alpina*, therefore, represents a suitable model for studying genomic adaptation in alpine perennial plants, for exploring adaptation to climate or other environmental factors, and for investigating the functions of candidate gene variants.

**Acknowledgments**

We are grateful to Marie Aline for help in *Arabis alpina* sampling. Many thanks to Mathieu Gautier and Michael Blum for discussions about this work.

## References


1. J. Hereford, A quantitative survey of local adaptation and fitness trade-offs, Am. Nat. 173 (2009) 579-588.
2. O. Savolainen, M. Lascoux, J. Merilä, Ecological genomics of local adaptation, Nat. Rev. Genet. 14 (2013) 807-820.
3. C. Weinig, B.E. Ewers, S.M. Welch, Ecological genomics and process modeling of local adaptation to climate, Curr. Opin. Plant. Biol. 18 (2014) 66-72.
4. S.D. Schoville, et al., Adaptative Genetic Variation on the Landscape: Methods and Cases, Annu. Rev. Ecol. Evol. Syst. 43 (2012) 23-43.
5. C.W. Ahrens, et al., The search for loci under selection: trends, biases and progress, Mol. Ecol. 27 (2018) 1342-1356.
6. S. Hoban, et al., Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions, Am. Nat. 188 (2016) 379-397.
7. S. Goodwin, J.D. McPherson, W.R. McCombie, Coming of age: ten years of next-generation sequencing technologies, Nat. Rev. Genet. 17 (2016) 333-351.
8. R.J. Elshire, et al., A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species, PLoS One 6 (2011) e19379.
9. N.A. Baird, et al., Rapid SNP discovery and genetic mapping using sequenced RAD markers, PLoS One 3 (2008) e3376.
10. A.P. Fuentes-Pardo, D.E. Ruzzante, Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations, Mol. Ecol. 26 (2017) 5369-5406.

11. C. Rellstab, et al., Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species, PLoS One 8 (2013) e80422.
12. Y. Zhu, et al., Empirical validation of pooled whole genome population re-sequencing in Drosophila melanogaster, PLoS One 7 (2012) e41901.
13. M. Ozerov, et al., Cost-effective genome-wide estimation of allele frequencies from pooled DNA in Atlantic salmon (Salmo salar L.), BMC Genomics 14 (2013) 12.
14. M.A. Fustier, et al., Signatures of local adaptation in lowland and highland teosintes from whole-genome sequencing of pooled samples, Mol. Ecol. 26 (2017) 2738-2756.
15. K.N. Stölting, et al., Genome-wide patterns of differentiation and spatially varying selection between postglacial recolonization lineages of Populus alba (Salicaceae), a widespread forest tree, New Phytol. 207 (2015) 723-34.
16. C. Körner, The use of 'altitude' in ecological research, Trends Ecol. Evol. 22 (2007) 569-574.
17. C. Körner, Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems, 2003, Springer, Heidelberg.
18. S. Lobréaux, S. Manel, C. Melodelima, Development of an Arabis alpina genomic contig sequence data set and application to single nucleotide polymorphisms discovery. Mol. Ecol. Resour. 14 (2014) 411-418.
19. E.M. Willing, et al., Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation, Nat. Plants 1 (2015) 14023.
20. W.B. Jiao, et al., Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data, Genome Res. 27 (2017) 778-786.

21. S. Bergonzi, et al., Mechanisms of age-dependent response to winter temperature in perennial flowering of Arabis alpina, Science 340 (2013). 1094-1097.
22. R. Wang, et al., PEP1 regulates perennial flowering in Arabis alpina, Nature 459 (2009) 423-427.
23. B.N. Poncet, et al., Tracking genes of ecological relevance using a genome scan in two independent regional population samples of Arabis alpina, Mol. Ecol. 19 (2010). 2896-2907.
24. D. Zulliger, E. Schnyder, F. Gugerli, Are adaptive loci transferable across genomes of related species? Outlier and environmental association analyses in Alpine Brassicaceae species, Mol. Ecol. 22 (2013) 1626-1639.
25. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (2009) 1754-1760.
26. C. Melodelima, S. Lobréaux, Complete Arabis alpina chloroplast genome sequence and insight into its polymorphism, Meta Gene 1 (2013) 65-75.
27. H. Li, et al., The Sequence Alignment/Map format and SAMtools, Bioinformatics 25 (2009) 2078-2079.
28. N. Takezaki, M. Nei, K. Tamura, POPTREEW: web version of POPTREE for constructing population trees from allele frequency data and computing some other quantities, Mol. Biol. Evol. 31 (2014) 1622-1624.
29. A.A. Popescu, K.T. Huber, E. Paradis, ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R, Bioinformatics 28 (2012) 1536-1537.
30. D. Chessel, A.B. Dufour, J. Thioulouse, The ade4 Package – I: One-Table Methods, R News 4 (2004) 5-10.
31. S. Lobréaux, C. Melodelima, Detection of genomic loci associated with environmental variables using generalized linear mixed models, Genomics 105 (2015) 69-75.

32. D. Bates, M. Mächler, B. Bolker,S. Walker, Fitting Linear Mixed-Effects Models Using lme4, J. Stat. Soft. 67 (2015) 1-48.
33. J.P. Shaffer, Multiple hypothesis-testing, Ann. Rev. of Psychol. 46 (1995) 561-584.
34. S. Joost, A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation, Mol. Ecol. 16 (2007) 3955-3969.
35. E. Frichot, E., et al., Testing for associations between loci and environmental gradients using latent factor mixed models, Mol. Biol. Evol. 30 (2013) 1687-1699.
36. O. Kürsteiner, I. Dupuis, C. Kuhlemeier, The pyruvate decarboxylase1 gene of Arabidopsis is required during anoxia but not other environmental stresses, Plant Physiol. 132 (2003) 968-978.
37. S.Y. Kim, et al., Exploring membrane-associated NAC transcription factors in Arabidopsis: implications for membrane biology in genome regulation, Nucleic Acids Res 35 (2007) 203-213.
38. P.J. Seo, Recent advances in plant membrane-bound transcription factor research: emphasis on intracellular movement, J. Integr. Plant Biol. 56 (2014) 334-342.
39. Y.N. Chen, E. Slabaugh, F. Brandizzi, Membrane-tethered transcription factors in Arabidopsis thaliana: novel regulators in stress response and development, Curr. Opin. Plant Biol. 11 (2008) 695-701.
40. P.J. Seo, S.G. Kim, C.M. Park, Membrane-bound transcription factors in plants, Trends Plant Sci. 13 (2008) 550-556.
41. P.J. Seo, et al., Cold activation of a plasma membrane-tethered NAC transcription factor induces a pathogen resistance response in Arabidopsis, Plant J. 61 (2010) 661-671.
42. H.K. Yoon, et al., Regulation of leaf senescence by NTL9-mediated osmotic stress signaling in Arabidopsis, Mol. Cells 25 (2008) 438-445.

43. M.J. Kim, et al., Controlled nuclear import of the transcription factor NTL6 reveals a cytoplasmic role of SnRK2.8 in the drought-stress response, Biochem J. 448 (2012) 353-363.

44. Y.S. Kim, et al., A membrane-bound NAC transcription factor regulates cell division in Arabidopsis, Plant Cell 18 (2006) 3132-3144.

45. S.G. Kim, et al., A membrane-bound NAC transcription factor NTL8 regulates gibberellic acid-mediated salt signaling in Arabidopsis seed germination, Plant J. 55 (2008) 77-88.

46. B.L. Orvar, et al., Early steps in cold sensing by plant cells: the role of actin cytoskeleton and membrane fluidity, Plant J. 23 (2000) 785-794.

47. P.J. Seo, et al., Proteolytic processing of an Arabidopsis membrane-bound NAC transcription factor is triggered by cold-induced changes in membrane fluidity, Biochem J. 427 (2010) 359-367.

48. M. Griffith, M.W. Yaish, Antifreeze proteins in overwintering plants: a tale of two activities, Trends Plant Sci. 9 (2004) 399-405.

49. C.S. Snider, et al., Role of ice nucleation and antifreeze activities in pathogenesis and growth of snow molds, Phytopathology 90 (2000) 354-361.

50. T. Nakajima, J. Abel, Development of resistance to Microdochium nivale in winter wheat during autumn and decline of the resistance under snow, Can J. Bot. 72 (1994) 1211-1215.

51. W.C. Hon, et al., Antifreeze proteins in winter rye are similar to pathogenesis-related proteins, Plant Physiol. 109 (1995) 879-889.

52. P. Toräng, et al., Large-scale adaptive differentiation in the alpine perennial herb Arabis alpina, New Phytol. 206 (2015) 459-470.

53. K. Tatjana, et al., Identification of genes involved in the response of leaves of Phaseolus vulgaris to drought stress, Molecular Breeding 21 (2008) 159:172.

54. J.M. Shockey, M.S. Fulda, J. Browse, Arabidopsis contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme a synthetases, Plant Physiol. 132 (2003) 1065-1076.

55. L. Rizhsky, et al., When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress, Plant Physiol. 134 (2004) 1683-1396.

56. A. Rohde, S. Kurup, and M. Holdsworth, ABI3 emerges from the seed, Trends Plant. Sci. 5 (2000) 418-419.

57. L. Niu, L., et al. Regulation of flowering time by the protein arginine methyltransferase AtPRMT10, EMBO Rep. 8 (2007) 1190-1195.

58. A. Lazaro, E. Obeng-Hinneh, M.C. Albani, Extended Vernalization Regulates Inflorescence Fate in Arabis alpina by stably silencing Perpetual Flowering1, Plant Physiol. 176 (2018) 2819-2833.

59. A.J. Palmer, A. Baker, S.P. Muench, The varied functions of aluminium-activated malate transporters-much more than aluminium resistance, Biochem. Soc. Trans. 44 (2016) 856-862.

60. S.G. Palusa, et al., Organ-specific, developmental, hormonal and stress regulation of expression of putative pectate lyase genes in Arabidopsis, New Phytol. 174 (2007) 537-550.

61. C. Valon, C., et al., Characterization of an Arabidopsis thaliana gene (TMKL1) encoding a putative transmembrane protein with an unusual kinase-like domain, Plant Mol. Biol. 23 (1993) 415-421.

62. A. Fournier-Level, et al., A map of local adaptation in Arabidopsis thaliana, Science 334 (2011) 86-89.

63. A.M. Hancock, et al., Adaptation to climate across the Arabidopsis thaliana genome, Science 334 (2011) 83-86.
64. J.R. Lasky, et al., Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate, Mol. Ecol. 21 (2012) 5512-5529.
65. Y. Zhen, M.C. Ungerer, Clinal variation in freezing tolerance among natural accessions of Arabidopsis thaliana, New Phytol. 177 (2008) 419-427.
66. A. Aydelotte, P. Diggle, Analysis of developmental preformation in the alpine herb Caltha leptosepala (Ranunculaceae), Am. J. Bot. 84 (1997) 1646.

## Figure legends

**Fig. 1.** Phylogenetic tree of the ten *Arabis alpina* populations studied based on SNP data. A phylogenetic tree was built using the neighbor-joining method from a set of 8,000 SNPs by randomly selecting 1,000 SNPs from each *Arabis alpina* chromosome. The three major groups were detected in the tree and are indicated as A, B, and C.

**Fig. 2.** The region containing *AaNAC062* gene is a major locus associated with the environmental climatic variable. A) Schematic representation of the chromosome-5 genomic region from position 11,521,406 to 11,546,695 bp. The different genes are indicated by an arrow from ATG to stop codons, with names according to the *Arabis alpina* genome annotations. Black boxes correspond to exons. Intergene sequences are represented as dashed lines, and their length is indicated.

**Fig. 3.** Partial alignment of NAC062 protein from *Arabis alpina* (AaG315210.t1), *Arabidopsis thaliana* NAC062 (AT3G49530.1), *Arabidopsis lyrata subsp lyrata* (XP_002877683.1) and *Capsella rubella* (XP_006291082.1) from amino acids 171 to 189. The genetic variation detected between *Arabis alpina* populations in the climatic gradient, and responsible for a non-synonymous mutation, is indicated by a star at position 181.

**Fig. 4.** Plot of the allelic frequency profile for a sample SNP from *AaNAC062* gene against the synthetic environmental variable used in the study. The upper part of the gradient at high elevation corresponds to positive values of the variable.

## Table 1

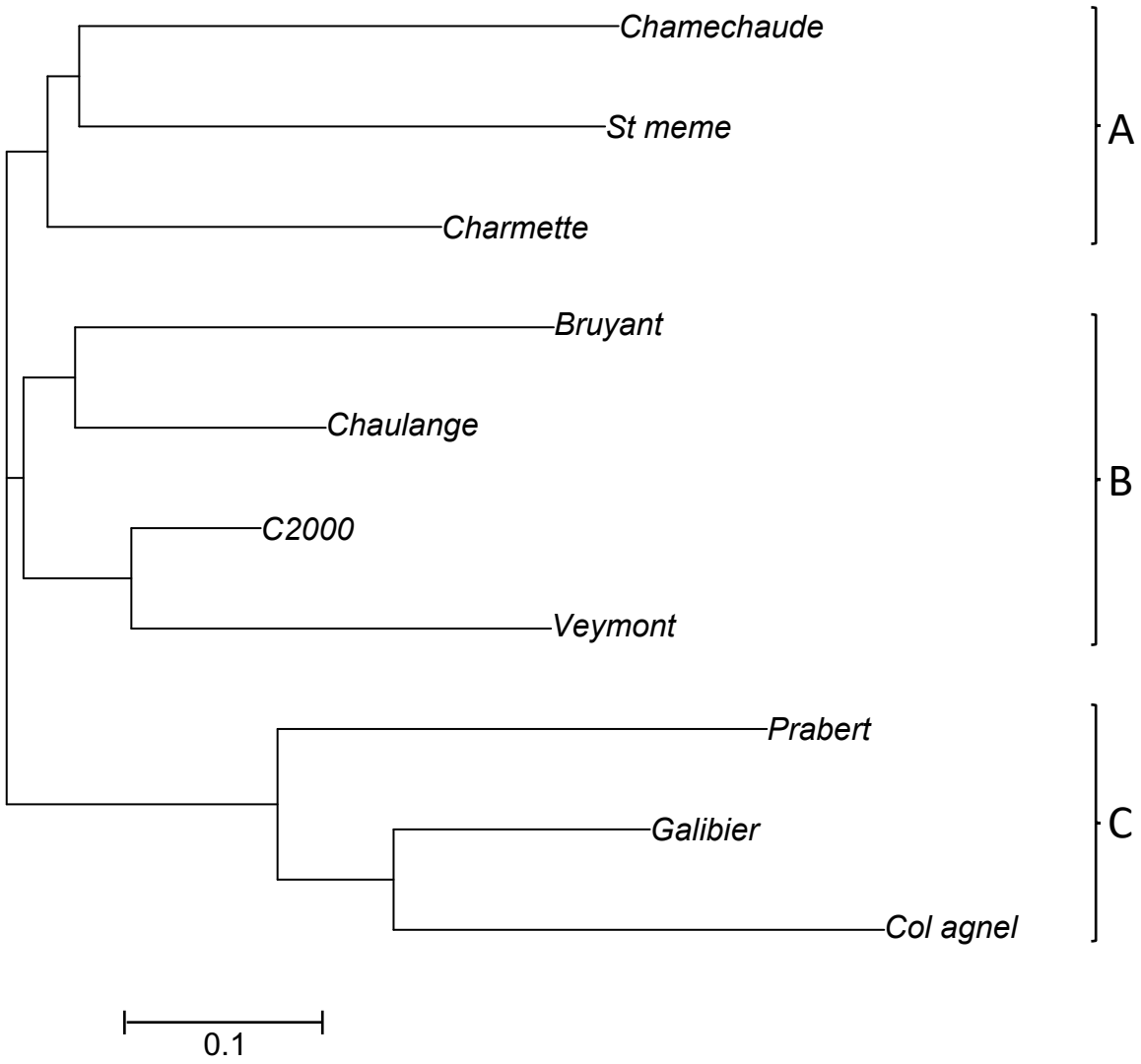
Sampling site data and synthetic climatic variable corresponding to axis 1 values from PCA of environmental variables

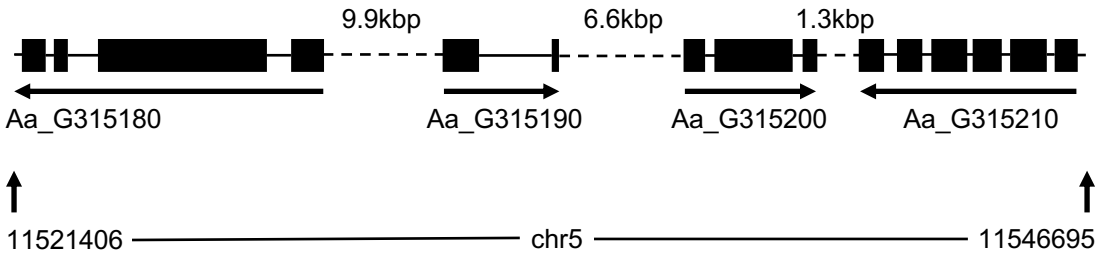
| Name | Latitude | Longitude | Elevation (m) | Synthetic variable |
|---|---|---|---|---|
| **St meme** | 45.39750 | 5.89125 | 860 | -2.4427 |
| **Bruyant** | 45.15069 | 5.61153 | 980 | -1.4517 |
| **Charmette** | 45.32818 | 5.73670 | 1364 | -0.4670 |
| **Prabert** | 45.21890 | 5.99395 | 1380 | -1.4145 |
| **Chaulange** | 45.07060 | 5.59411 | 1545 | -1.0925 |
| **Chamechaude** | 45.28571 | 5.78504 | 1914 | 0.6757 |
| **Veymont** | 44.88069 | 5.52211 | 1950 | 0.4184 |
| **C2000** | 45.01968 | 5.57005 | 1951 | 0.1420 |
| **Galibier** | 45.06058 | 6.40485 | 2536 | 2.9396 |
| **Col agnel** | 44.68568 | 6.98280 | 2781 | 2.6927 |

## Table 2

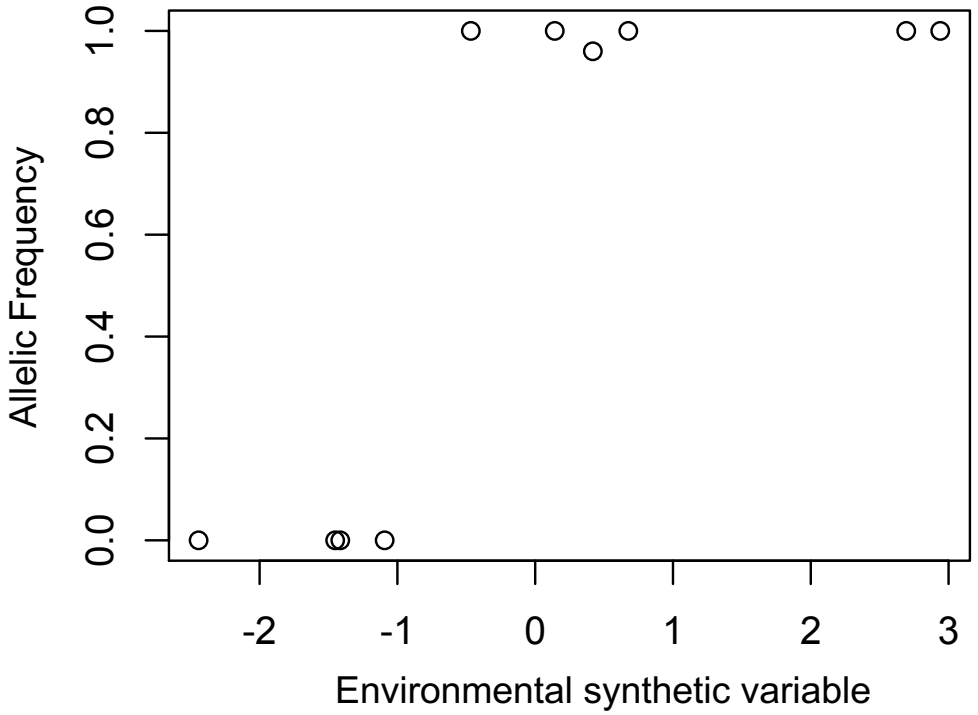
*Arabis alpina* candidate genes within genomic regions significantly associated with the synthetic environmental variable. Potential candidate genes within these regions are indicated together with their *Arabidopsis thaliana* ortholog when available.

| Chr | localization | Aa_ID | Function | Arabidopsis ortholog |
|---|---|---|---|---|
| chr1 | 1132833-1137852 | Aa_G6950 | uncharacterized protein | AT1G02990 |
| chr1 | 4989967-4992419 | Aa_G368310 | aluminum activated malate transporter | AT1G08440 |
| chr2 | 35950629-35952085 | Aa_G797250 | hxxxd-type acyl-transferase-like protein | AT1G78990 |
| chr3 | 1816190-1819332 | Aa_G185650 | tcp-1 cpn60 chaperonin family protein | AT3G03960 |
| chr3 | 17087273-17089595 | Aa_G284660 | uncharacterized protein | AT3G24750 |
| chr3 | 17108584-17111085 | Aa_G284690 | pectate lyase family protein | AT3G24670 |
| chr3 | 17111534-17113646 | Aa_G284700 | kinase-like protein tmkl1-like | AT3G24660 |
| chr3 | 17136875-17137113 | Aa_G385880 | leucine-rich repeat-containing protein | - |
| chr3 | 17140832-17143771 | Aa_G385870 | abscisic acid-insensitive protein 3 | AT3G24650 |
| chr5 | 11544865-11546695 | Aa_G315210 | NAC062 transcription factor | AT3G49530 |
| chr5 | 12159593-12705643 | Aa_G53000 | Kinesin family-like protein | - |
| chr6 | 2380152-2382801 | Aa_G89730 | pfkb-like carbohydrate kinase protein | AT5G43910 |
| chr6 | 30161491-30162716 | Aa_G245490 | vesicle transport v-snare | AT5G39510 |
| chr8 | 35675667-35677688 | Aa_G54570 | protein arginine n-methyltransferase | AT1G04870 |
| chr8 | 45693680-45697947 | Aa_G97990 | acyl-activating enzyme 17 | AT5G23050 |


Chamechaude
St meme
Charmette
Bruyant
Chaulange
C2000
Veymont
Prabert
Galibier
Col agnel
A
B
C
0.1


9.9kbp
6.6kbp
1.3kbp
Aa_G315180
Aa_G315190
Aa_G315200
Aa_G315210
11521406
chr5
11546695

```
                                *
Aa_NAC062       171  EEDSKSDEAEPEPAGSSPT 189
                     |||||||||:| ||| ||||
At_NAC062       171  EEDSKSDEVE-EPAVSSPT 188
```


Allelic Frequency
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
3
Environmental synthetic variable