bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Phenotypic and genomic differentiation of *Arabidopsis thaliana* along altitudinal gradients in the North Italian alps

Torsten Günther[$,‡,§], Christian Lampei[$,‡], Ivan Barilar[$], and Karl J. Schmid[$,*]

July 22, 2015

[$] Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany
[§]Department of Evolutionary Biology, EBC, Uppsala University, Uppsala, Sweden.
[*]Corresponding author: E-mail: karl.schmid@uni-hohenheim.de.
[‡]These authors contributed equally to this work.

Abstract

Altitudinal gradients represent short-range clines of environmental parameters like temperature, radiation, seasonality and pathogen abundance, which allows to study the footprints of natural selection in geographically close populations. We investigated phenotypic variation for frost resistance and light response in five *Arabidopsis thaliana* populations ranging from 580 to 2,350 meters altitude at two different valleys in the North Italian Alps. All populations were resequenced as pools and we used a Bayesian method to detect correlations between allele frequencies and altitude while accounting for sampling, pooled sequencing and the expected amount of shared drift among populations. The among population variation to frost resistance was not correlated with altitude. An anthocyanin deficiency causing a high leaf mortality was present in the highest population, which may be non-adaptive and potentially deleterious phenotypic variation. The genomic analysis revealed that the two high-altitude populations are more closely related than the geographically close low-altitude populations. A correlation of genetic variation with altitude revealed an enrichment of highly differentiated SNPs located in genes that are associated with biological processes like response to stress and light. We further identified regions with long blocks of presence absence variation suggesting a sweep-like pattern across populations. Our analysis indicate a complex interplay of local adaptation and a demographic history that was influenced by glaciation cycles and/or rapid seed dispersal by animals or other forces.

Introduction

Local adaptation results from interactions of populations with biotic and abiotic environmental conditions. Numerous phenotypic traits that contribute to adaptation show therefore clinal variation along environmental gradients (Huxley, 1938). Phenotypic clines have been investigated for many years (Mayr, 1942) because predictions regarding patterns and processes of natural selection can be tested in a rigorous fashion (Etterson & Shaw, 2001). In annual plants, an

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

example of a well studied cline is the change of flowering time with latitude gradients (Stinchcombe *et al*, 2004). The timing of the transition from the vegetative to the reproductive stage is strongly related to fitness in annual plants and therefore depends on day length and temperature.

In mountaineous environments, altitude is one of the strongest environmental gradients because abiotic parameters like temperature, radiation intensity, atmospheric pressure, and partial pressure of gases simultaneously change over short distances (Körner, 2007). The biotic environment includes competitors, herbivores, pathogens or symbionts whose occurrence differs with altitude as well. The combination of these factors results in different ecological niches at different altitudes, which likely exert a strong selection pressure for local adaptation. In addition to external factors, the intrinsic genetic potential for adaptation is affected by altitude since the absolute area available for plant growth tends to rapidly decrease with altitude (Körner, 2007) causing small population sizes and increasing levels of genetic drift and inbreeding at high altitudes (Thiel-Egenter *et al*, 2011; Manel *et al*, 2012).

Another critical factor in adaptation to different altitudes is the age of local populations. Long-term changes in temperature and other environmental parameters caused by glacial cycles contributed to large fluctuations in the altitudinal distribution of species, which retreated from high altitudes during glacial maxima and expanded during interglacial periods. However, numerous stable glacial refugial populations on mountain tops (nunataks) existed in the European Alps (Schönswetter *et al*, 2005) and other mountain ranges. In such populations, selection and genetic drift likely contributed to the rapid genetic and phenotypic divergence from other populations of the same species, and frequently led to the formation of endemic sibling species restricted to high altitudes.

Arabidopsis thaliana shows a wide distribution and local adaptation within the Northern hemisphere for many morphological, life history and other fitness-related traits (reviewed by Koornneef *et al*, 2004; Bergelson & Roux, 2010). Genomic surveys suggested adaptation to large-scale environmental gradients in central Europe (Clark *et al*, 2007; Hancock *et al*, 2011; Horton *et al*, 2012), the Iberian Peninsula (Méndez-Vigo *et al*, 2011) and Scandinavia (Long *et al*, 2013; Huber *et al*, 2014). Adaptation to different altitudes can also be studied in *A. thaliana* since it occurs from the sea level up to 4,250 m altitude (Al-Shehbaz & O'Kane, 2002). High altitude adaptation was observed in populations in the eastern Pyrenees and included traits like fecundity, phenology and biomass allocation, suggesting selection for higher vigour in high altitude plants (Montesinos-Navarro *et al*, 2011), selection for alpine dwarfism in the Swiss Alps (Luo *et al*, 2015), or a constitutive protection against UV-B radiation protection in Shakdara, Central Asia at 3,063 m a.s.l (Biswas & Jansen, 2012).

We investigated the roles of adaptive evolution and demographic history in the distribution of genetic variation and putative adaptive phenotypic traits along an altitudinal gradient in the European Alps. We sampled five *A. thaliana* populations at different altitudes up to 2,300 m in the South Tyrolia and Trento provinces of Italy. Four populations represent two pairs of high and low altitudes in two different valleys, and and a fifth population was located at equal

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

distance to both gradients. In these populations, we characterized phenotypic differentiation in frost resistance and response to light and UV-B stress because these environmental parameters are strongly associated with altitude (Blumthaler *et al*, 1997; Körner, 2007). Accessions from the Northern Italian Alps were previously shown to be genetically distinct from other regions and to harbour reduced genetic diversity (Cao *et al*, 2011). We reasoned that high altitude adaptation resulted in a strong signal of genetic differentiation relative to the genome-wide background, which should allow to identify selected genomic regions. We sequenced pools of individuals from each of the five populations, because pool sequencing is a cost-efficient means to analyse diversity in populations to infer demography and selection, and was previously applied in several plant and animal species (e.g. Turner *et al*, 2010; He *et al*, 2011; Kolaczkowski *et al*, 2011; Fabian *et al*, 2012; Lamichhaney *et al*, 2012; Orozco-terWengel *et al*, 2012; Fischer *et al*, 2013). Although sequencing of pools comes at the cost of noisy estimates and a loss of linkage information (Futschik & Schlötterer, 2010; Zhu *et al*, 2012), suitable frameworks that account for the special properties of pooled sequence data are available (Futschik & Schlötterer, 2010; Kofler *et al*, 2011b,a; Boitard *et al*, 2012; Günther & Coop, 2013). We investigated the demographic history by Approximate Bayesian Computation (ABC, Beaumont *et al*, 2002), and correlated allele frequencies with altitude to identify genomic regions that are significantly differentiated between high and low altitude populations because of local adaptation. We observed a complex pattern of genomic and phenotypic diversity suggesting that adaptation to high altitudes may affect genes involved in light responses and soil conditions as well as other stress factors, but also indicate that there is no simple pattern of genomic and phenotypic differentiation with altitude.

Materials and Methods

Plant material

Seeds were sampled at five locations in the alps of South Tyrolia and Trento provinces to represent two pairs of high and low altitudes, respectively, in different valleys and mountain ranges but in close geographic proximity (Table 1 and Figure 1). All populations were located on steep slopes and exposed to the South (Coordinates: Vioz/Coro 46°22'57"N, 10°39'35"E; Terz 46°21'50"N, 10°55'13"E; Finail 46°44'31"N, 10°49'5"E; Juval 46°38'55"N, 10°58'27.85"E; Laatsch 46°40'18.00"N, 10°30'54.00"E). Two low altitude populations (Laatsch and Juval) are located on scree which was partly overgrown in Juval, but still moving in Laatsch. These two sites were the least disturbed of the five sites. The third valley population (Terz) is located on marginal soils in dry stone walls in the vicinity of a wine yard and was highly disturbed. The two high-altitude populations occur at South exposed rocky outcrops that serve as resting places of mountain goats, and therefore are strongly disturbed habitats with a high supply of nitrogen and phosphorous. Due to different seasonality, we visited the low altitude populations in early May and the high altitude populations in mid-July. All locations were sampled in a spatially balanced design to include the maximal diversity present at the site. Plants were

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1: Summary of pooled sequencing libraries.

			PoolSeq summary statistics			
Population	Individuals sequenced	Altitude [m]	Median coverage of nDNA	Percent genome covered	Nucleotide diversity, π (SE)	Tajima's D (SE)
Juval	29	577 - 657	32x	94.3	0.00305 (0.00003)	0.0352 (0.0063)
Terz	25	810 - 945	31x	95.8	0.00439 (0.00003)	-0.0107 (0.004)
Laatsch	29	970 - 1010	19x	94.5	0.00190 (0.00003)	-0.1912 (0.0084)
Finail	29	2250 - 2270	32x	94.2	0.00190 (0.00002)	-0.7567 (0.0046)
Vioz/Coro	19	2210 - 2355	50x	93.3	0.00166 (0.00002)	-0.7122 (0.0044)

cultivated for DNA extraction on standard gardening soil (Einheitserde ED 63T). Forty day old whole plants were freeze dried in a Christ Alpha 14 (Martin Christ Gefriertrocknungsanlagen GmbH, Osterode am Harz, Germany) freeze dryer before DNA extraction.

Phenotypic analysis of spring freezing tolerance and response to light stress and UV-B radiation

Freezing tolerance We investigated the plants for differences in freezing damage on the population level under two temperature regimes with plants from all populations grown in the greenhouse by adapting the method of Martin *et al* (2010). Eight random accessions per population were tested at a minimum temperature of -15°C, and five random accessions per population at -7°C minimum temperature. We raised the plants under long day condition (16 h light/8 h dark) at 23° C and adapted them to the cold for one day at 4° C at the eight leaf stage. Whole plants were sampled, washed with deionized water and wrapped in aluminum foil after removal of residual water.

We then put the samples into a steel vacuum Dewar bottle which was placed in a styrofoam box (14 l volume and 5 cm wall thickness). Two temperature loggers (3M™Temperature Logger TL20, 3M Deutschland GmbH, Neuss) were added to the Dewar bottle above and below the samples to monitor the realized treatment temperature. To ensure a slow temperature decrease, we added a PET-bottle with 1 l of saturated salt water (20% NaCl solution) to the styrofoam box. The box was kept at -20° C for 17 and 7 hours, respectively. Slow defrosting was ensured at 4°C for approximately 20 h. Samples were put into 10 ml of deionized water in sterile 15 ml centrifuge tubes and left at room temperature overnight to allow cytosol leakage of damaged cells. We then measured electric conductivity with a handheld conductivity meter (WTW LF 90, KLE1 sensor; measure 1). To obtain conductivity at 100% cell damage, we autoclaved all samples (Systec DB-23 benchtop autoclave) and measured conductivity again (measure 2). Freezing damage of each sample was calculated as the proportion of measure 1 over measure 2.

Light and UV-B stress We further characterized the response to different light conditions. Random individuals of each population were cultivated on standard gardening soil (Einheitserde ED 63T) at 23°C in a climate cabinet (GroBank BB-XXL^{3+}, CLF Plant Climatics, Emersacker, Germany) under three light treatments under long day conditions (16/8 light/dark): (1) A

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

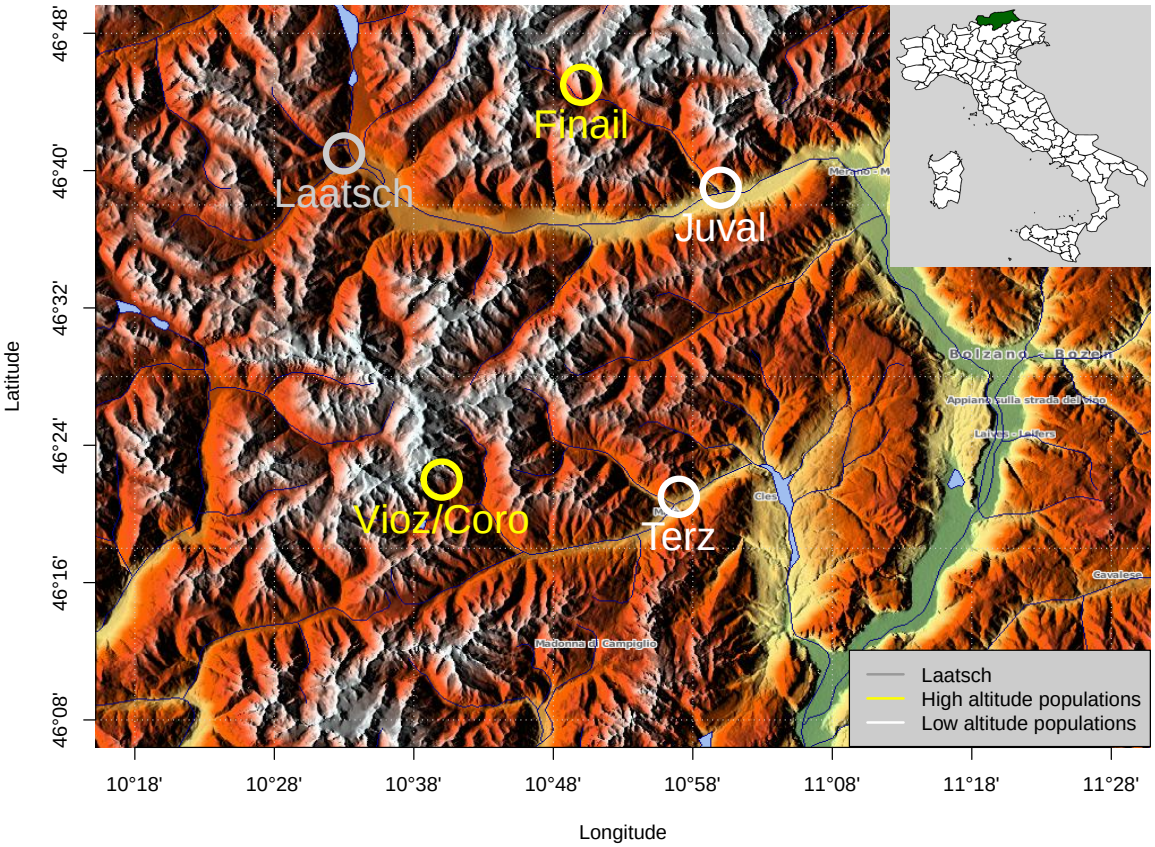

Figure 1: Geographic map showing the location of the five populations included in this study. Map source: `http://www.maps-for-free.com/`

control low light treatment, (2) a low light with UV-B treatment and (3) a light-stress treatment without UV-B. The low light treatment consisted of 168 μmol m^{-2} s^{-1} fluorescent light (Philips Master TL-D 58W/840 Reflex), and the UV-B treatment used the same fluorescent setting with additional UV-B radiation exposure three times per day for one hour at variable intensities (6:00−7:00 am 0.01 mWcm^{-2}, 11:00−12:00 am 0.027 mWcm^{-2} and 4:00-5:00 pm 0.027 mWcm^{-2}). In the light-stress treatment, we exposed plants to fluorescent light with an intensity of 470 μmol m^{-2} s^{-1}, which is more than twice the optimal light intensity for the Col-0 ecotype (200 μmol m^{-2} s^{-1}). In the UV-B treatment, fluctuating UV-B levels were judged to be more realistic than stable UV-B exposure because frequent cloud cover reduces the total annual amount of light and UV-B radiation but fluctuations in light intensity increase strongly with higher altitude (Körner, 2007). The leaf color of each plant was used as stress indicator because light or UV-B stressed plants show increased anthocyanin production which causes a red leaf colour (Chalker-Scott, 1999). All plants were photographed with a digital camera (Samsung NX 10) at high resolution from the top and customized macros of the ImageJ distribution Fiji (Schindelin *et al*, 2012) were used to measure the red leaf area, the total leaf area and the

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

area of dead plant material.

Statistical analysis of phenotypic data To identify differences in freezing damage between populations, we fitted a generalized linear model (GLM) with freezing damage odds as dependent variable and a population factor and a test-condition factor and their interaction as independent variables based on a quasibinomial error distribution (logit link) as implemented in the R package stats (DevelopmentCoreTeam, 2014). Although statistical theory suggests to use a binomial error distribution to analyse the proportion of read and dead tissue after the light treatment, we used a general least squares model (GLS) assuming a normal distributed error because the residual distribution did not deviate from a normal distribution (Shapiro-Wilk test, $W = 0.96$, $p = 0.19$) and only the GLS model allows to correct for the strong heteroscedasticity in the data. Using the nlme package in R (Pinheiro *et al*, 2014) we fitted GLS models with the red proportion or the dead proportion as dependent variable and population, treatment and their interaction as independent factors. We allowed for different variances for each level of population and treatment using the *varComb* function with two *varIdent* functions. To identify differences between treatments in both experiments, we performed *post-hoc* comparisons with the *glht* function from the R package multcomp and corrected for the false discovery rate (FDR; Benjamini & Hochberg, 1995).

Sequencing and population genetic analysis

DNA sequencing We ground dry whole plants with a Retsch MM400 centrifugal mill (Retsch GmbH, Haan, Germany) and extracted the DNA with a standard CTAB maxi prep protocol. DNA quality and concentration were determined on a 0.8% agarose gel and additionally with the Qubit 2.0 Fluorometer. For each of the five populations, we prepared a pooled DNA sample (5 μg) by mixing of equimolar amounts of DNA from all individuals. Each DNA pool was then converted into a barcode-tagged genomic sequencing library using the standard Illumina protocol. The five tagged libraries were pooled and sequenced together on a single lane of an Illumina HiSeq 2000 sequencer by GATC Biotech (Konstanz, Germany). We then trimmed the 101 bp long paired-end reads to consecutive runs of base calls with a quality ≥ 25 and filtered for a minimum read length of 35 bp using SolexaQA (Cox *et al*, 2010). The trimmed high-quality reads were mapped against the *A. thaliana* TAIR10 reference genome (Lamesch *et al*, 2012) using BWA version 0.6.2 (Li & Durbin, 2009) using the non-default parameters `-n 0.05` (mismatch rate) and disabled seeding of reads. The BWA module sampe mapped paired-end reads to a single SAM file and samtools version 0.1.18 (Li *et al*, 2009) filtered the resulting alignments to include only properly paired reads and a minimum mapping quality of 20. For all downstream analyses, we converted the data to pileup files with samtools mpileup. We then identified SNPs as polymorphic sites of at least 12x coverage in at least one pool with a minimum of two reads supporting a non-reference allele and polarized them with the close relative *Arabidopsis lyrata* (Hu *et al*, 2011).

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Population genomic analysis We further analysed the Pileup files with Popoolation version 1.2.2 (Kofler *et al*, 2011a) to calculate nucleotide diversity, π, Wattersons θ and Tajima's D separately for each population. First, each position was independently subsampled to create a uniform 18x coverage over all populations by random drawing base calls from the pileup file without replacement. Then, π, θ and Tajima's D were calculated separately for each population using sliding windows with window and step sizes of 2,000 bp each. Only sites with a coverage of at least 10 and maximally 100 were included in the next analysis steps.

We annotated SNPs with SnpEff version 3.1 (Cingolani *et al*, 2012) based on the TAIR10 annotation (Lamesch *et al*, 2012) and assigned them to one of the following classes: intergenic, 5' UTR, 3' UTR, intron, synonymous coding, non-synonymous coding, synonymous stop, stop lost, stop gained and start lost. SNPs belonging to different groups (e.g. non-synonymous coding and stop gained) were assigned to the stronger effect category (e.g., stop gained). We further annotated regulatory SNPs based on known and predicted transcription factor binding sites from AGRIS (Yilmaz *et al*, 2011). The effect of non-synonymous SNPs on protein function (Günther & Schmid, 2010) was determined with the MAPP program (Stone & Sidow, 2005), which classifies nonsynonymous SNPs into deleterious or tolerated amino acid substitutions based on a collection of homologous proteins obtained with PSI-BLAST searches (Altschul *et al*, 1997) of all *A. thaliana* proteins against Uniprot (Wu *et al*, 2006). We calculated a phylogenetic tree for these proteins with semphy (Friedman *et al*, 2002) and used the multiple protein sequence alignment and the tree as input for MAPP (Stone & Sidow, 2005).

Identification of strongly differentiated genes To identify strongly differentiated genes, we correlated altitude with allele frequencies within populations with Bayenv2.0 because it models the sampling noise inherent to pooled sequencing and accounts for a shared history of populations (Günther & Coop, 2013). To characterize this relationship, we estimated a covariance matrix with Bayenv2.0 from a subsample of 10,000 high confidence SNPs (i.e. SNPs with a coverage between 30 and 100 in each of the five populations) based on 1,000,000 MCMC iterations. To ensure convergence we estimated a second matrix from another subsample of 10,000 SNPs. For the Bayenv2.0 analysis, we included only SNPs for which both alleles segregated in at least two populations and the per population sequence coverage was between 15 and 100, which resulted in a total of 400,231 SNPs. Using these data we estimated the genotype-environment correlation statistics Z and $\rho(X, Y')$ (Günther & Coop, 2013) between allele frequencies and altitude per population (normalized to mean 0 and standard deviation 1) estimated during 200,000 MCMC iterations.

To further analyse the functional annotation of highly differentiated genes, we searched for an enrichment of different annotation categories obtained from AraPath (Lai *et al*, 2012), which comprises gene ontology (GO, Ashburner *et al*, 2000), KEGG (Kanehisa *et al*, 2012), AraCyc (Mueller *et al*, 2003) and plant ontology (PO, Cooper *et al*, 2013) annotations for all *A. thaliana* genes, and additionally includes gene sets obtained from extensive literature research (Lai *et al*, 2012). To avoid a biased enrichment analysis, we tested enrichment of annotations using Gowinda (Kofler *et al*, 2012), which accounts for gene length using a permutation ap-

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

proach. We conducted 100,000 permutations and included all annotation categories of at least two genes while all SNPs in a gene and within 2,000 bp distance of a transcript were assigned to the respective gene.

Demographic history We investigated the genetic relationship of populations with Neighbour-joining trees (Saitou & Nei, 1987) based on the matrix of average pairwise F_{ST} values calculated by PoPoolation2, and from a correlation matrix of allele frequencies derived from the covariance matrix estimated by Bayenv2.0. The trees were plotted with the R package `ape` (Paradis *et al*, 2004). Additionally, we reconstruced the historical relationship among the five populations using their current genome-wide allele frequencies with TreeMix (Pickrell & Pritchard, 2012). We carried out a run which included all non-singleton SNPs segregating in all five populations with a minimum coverage of 20 per population (1,071,527 SNPs). We bootstrapped each run 1,000 times and carried out a four-population test which tests whether the relationship of the two pairs of high and low populations can be explained by any of the possible tree topologies (Reich *et al*, 2009; Patterson *et al*, 2012) to verify the TreeMix results.

We estimated divergence times between pairs of high and low altitude populations by simulating models of demographic history with msABC (Pavlidis *et al*, 2010) followed by an Approximate Bayesian Computation (ABC) analysis with the `abc` R package (Csilléry *et al*, 2012). We investigated four different models under a population-split scenario to estimate divergence times between low and high altitude populations (Figure S3). All models assumed a common ancestral population with a founding bottleneck leading to the high altitude population, but models differed in their post-split growth parameters of the high altitude populations: (i) the no-growth model assumed a constant population size after the split, (ii) the step-growth model assumed an instantaneous population size change some time after the split, (iii) the growth model assumed an exponential growth of the population after the split where the growth rate was calculated based on a previously drawn current population size (N_0; Table S3) and the time of the split, and (iv) the exponential model assumed exponential growth of the population size after the split. We simulated each model 200,000 times with and without bidirectional migration and included them in a model selection step with cross-validation based on selected summary statistics (see below) to select the best-fitting model. Posterior model probabilities were calculated with the *postpr* function of the R `abc` package using the *mnlogistic* method and a tolerance rate of 0.005. Using the best fitting model, we inferred the posterior distribution of the population split time.

In the simulations we assumed uniform priors for parameters (Table S3), and obtained mutation and recombination rates used in simulations from empirical studies (Kim *et al*, 2007; Ossowski *et al*, 2010; Cao *et al*, 2011) or based them on plausibility arguments (size of the high altitude populations, population growth parameter, time of the split). The ABC analysis included the following summary statistics: mean and variance of Tajima's D and π, mean of Watterson's θ and average number of segregating sites for each of the populations; F_{ST} values and numbers of shared, fixed and private SNPs, which were calculated for each population pair. We calculated summary statistics from 2 kb windows and randomly sampled 1,000 windows from

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the sequence data. Posterior parameter distributions were estimated by comparing observed and simulated summary statistics using the *abc* function of the R abc package with the *rejection* method and a tolerance rate of 0.01. From the posterior distributions of the best models, we calculated the Highest Posterior Density Interval (HPDI 90) of the estimated mean of the population sizes and the time of their split using the LaplacesDemon R package (Hall, 2013). The HPDI 90 measurement gives the most accurate 90% credible interval (Bayesian equivalent of confidence intervals) of the posterior distributions.

Results

Phenotypic differences between high and low altitude populations In the test for freezing resistance we first decreased the temperature to -15°C with a shoulder at around -7°C, which marks the begin of major tissue freezing (Figure 2 B). In the second experiment, temperatures were lowered to only -7° C and no shoulder was observed in the temperature curves (Figure 2 D). The two freezing treatments caused strong differences in the degree of tissue damage (Figure 2 A+C, $F_{1/59}$=128, p<0.001) with a population median from 0.8 to 0.9 in the first test and from 0.1 to 0.6 in the second test. Based on the GLM there was no significant temperature × population interaction ($F_{4/55}$=1.42, p=0.24), but differences in freezing damage between populations ($F_{4/59}$=7.16, p<0.001). A post-hoc multiple comparison test (α=0.05, FDR corrected following (Benjamini & Hochberg, 1995)) combined the populations Juval, Laatsch and Vioz/Coro into a homogeneous subset with stronger leaf damage and the populations Terz and Finail in a subset with higher frost resistance (Figure 2). The differentiation suggests adaptation to local temperature regimes, but since both subsets include at least one high and one low altitude population and originated from different drainage systems, frost resistance in these populations is not strongly correlated with altitude nor with physico-geographic proximity.

As second phenotypic trait we investigated response to different types and strengths of light irradiation measured as red coloration of leaves, which is caused by anthocyanin accumulation and represents a typical stress reaction of plants (Park *et al*, 2007). The population × treatment (light condition) interaction was highly significant ($F_{8/28}$=42, p<0.001) indicating differential population response to light stress. A post-hoc multiple comparison test (α=0.01, FDR corrected following Benjamini & Hochberg (1995)) on the GLS model indicated that plants from four of the five populations developed more red tissue under light stress than in the other light conditions (Figure 3 A). Conversely, plants of the highest population (Vioz/Coro) were green in all treatments (Figure 3 A, D) but developed more dead tissue under the high light treatment in comparison to the low light and UV-B treatments (Figure 3 B, D, population × treatment interaction $F_{8/28}$=14, p<0.001).

An increased sensitivity to UV-B damage is expected in anthocyanin-deficient plants if the flavonoid pathway is affected simultaneously (Chalker-Scott, 1999). To test whether plants from high-altitude Vioz/Coro site showed an increased UV-B sensitivity, we exposed them to a gradient of increasing UV-B dosages over a period of three days and evaluated the proportion of dead tissue (Supplementary Methods). In contrast to the Finail population, which exhibited a

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

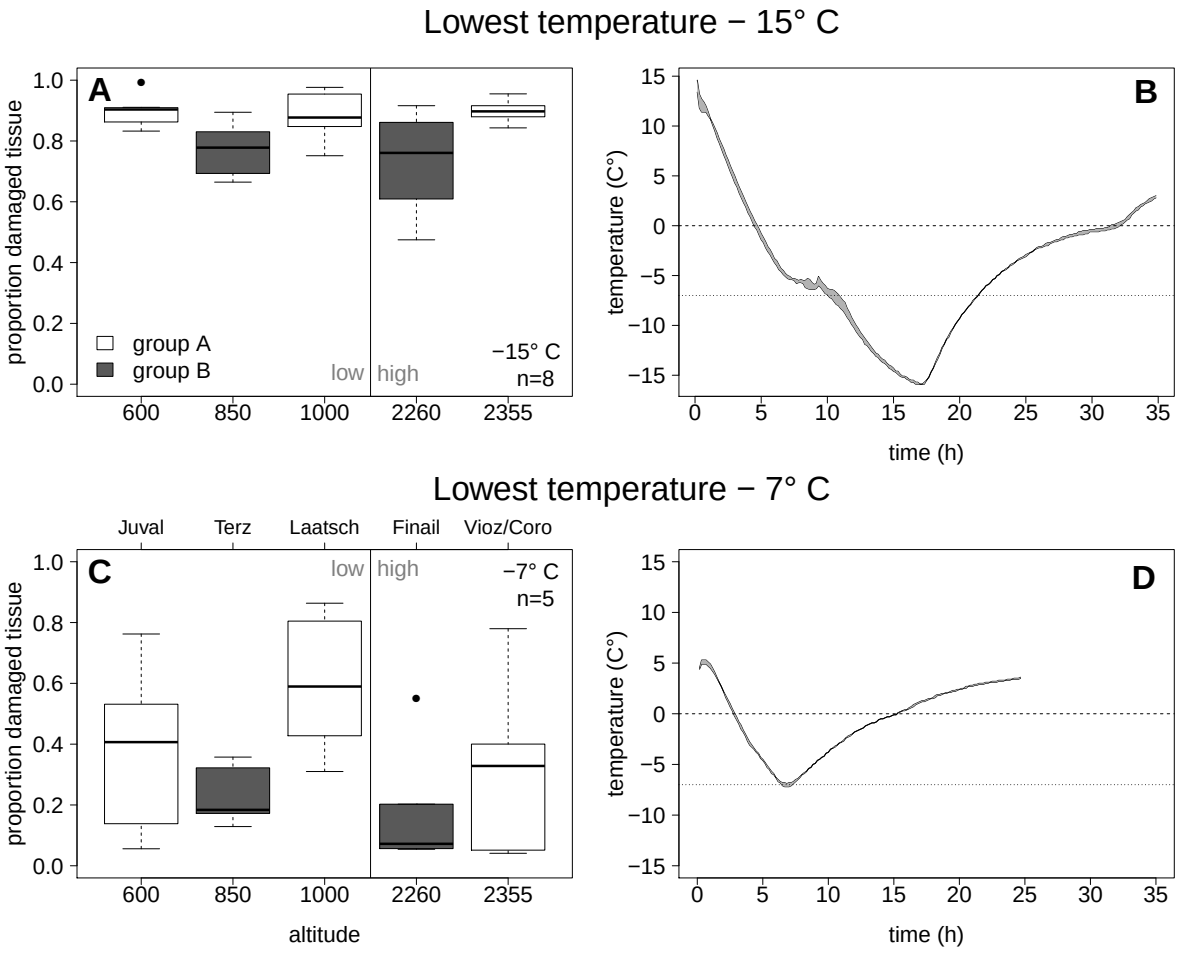

Figure 2: Freezing damage measured as cytosol leakage of destroyed tissue after freezing down to -15° C (A) and -7° C (C) as a fraction of cytosol leakage from completely destroyed tissue ($n = 8$). The vertical line divides the figure into the low altitude and the high altitude populations. Group A and B indicate homogeneous population subsets ($\alpha = 0.05$) after correction for false discovery rate. (B) and (D) show the logged temperature as area between the measurements of two data loggers positioned above and below the probes.

superior UV-B resistance compared to the low altitude populations, plants from the Vioz/Coro population developed an increased proportion of dead tissue with higher UV-B dosage (binomial GLM; population × UV-B dosage $F_{4/186}$=22, p<0.001; Figure S1, Table S1).

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

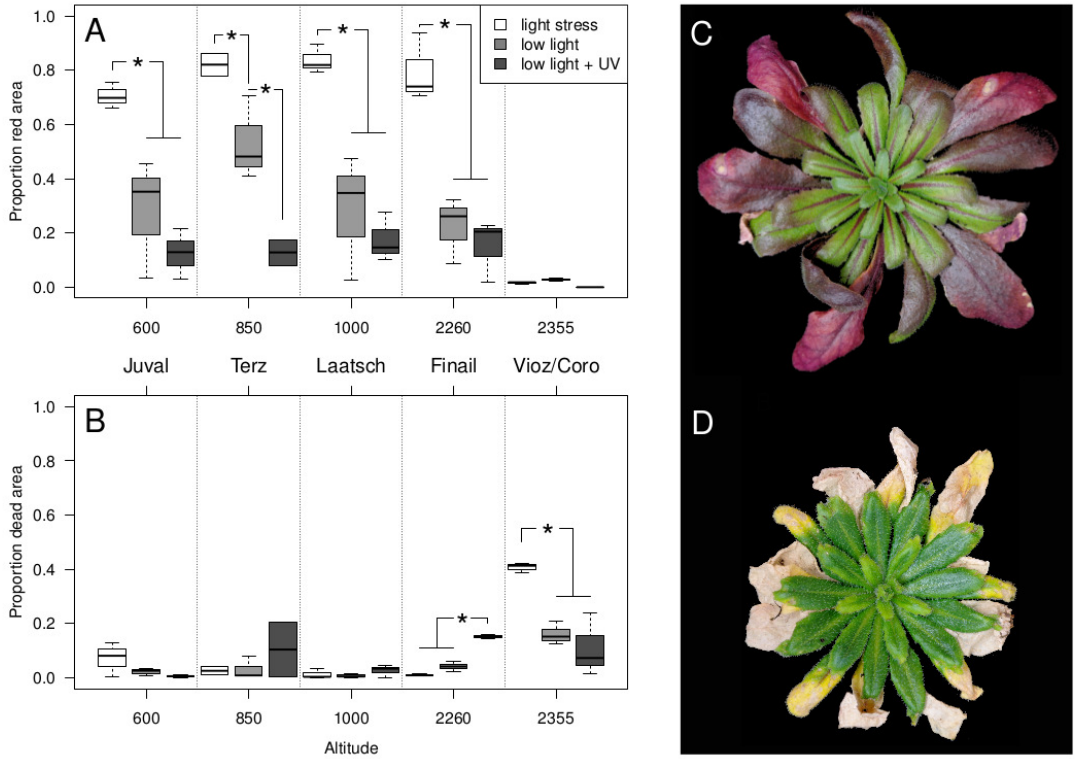

Figure 3: Phenotypic effects of 20 days of light stress treatment. A) Proportion of red tissue in light-stressed plants along the altitude gradient. B) Proportion of dead leaf tissue in light-stressed plants along the altitude gradient. Significant treatment differences at $p < 0.001$ or smaller (corrected for false discovery rate) are indicated with a star. C) A representative plant for the light stress treatment from the Finail population. D) A representative plant for the light stress treatment from the Vioz/Coro population.

Genetic diversity within and genetic relationship between populations Pool sequencing of all five populations and subsequent mapping to the Col-0 genome resulted in a 19- to 50-fold median coverage of the nuclear chromosomes per population (Table 1). In each pool, about 95% of the reference genome sequence positions were sequenced at least once and a total of 3,075,350 SNPs were called. We first calculated nucleotide diversity, π, and Tajima's D by accounting for pooling (Futschik & Schlötterer, 2010; Kofler *et al*, 2011a). As observed before (Clark *et al*, 2007; Cao *et al*, 2011), diversity was elevated in pericentromeric regions relative to chromosome arms (Figure 4A). The genome-wide average diversity of low altitude populations was high, while average Tajima's D was close to 0 (Figure 4B, Table 1). In contrast, high altitude populations show a reduced nucleotide diversity and negative Tajima's D values.

Allele frequency distributions were similar across populations with some interesting differences (Figure S4). All populations showed a large number of rare alleles and a substantial number of high-frequency or fixed derived alleles. The proportion of SNPs segregating at intermediate frequencies differed between populations (Table 1, Figure S4). SNPs with a strong

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

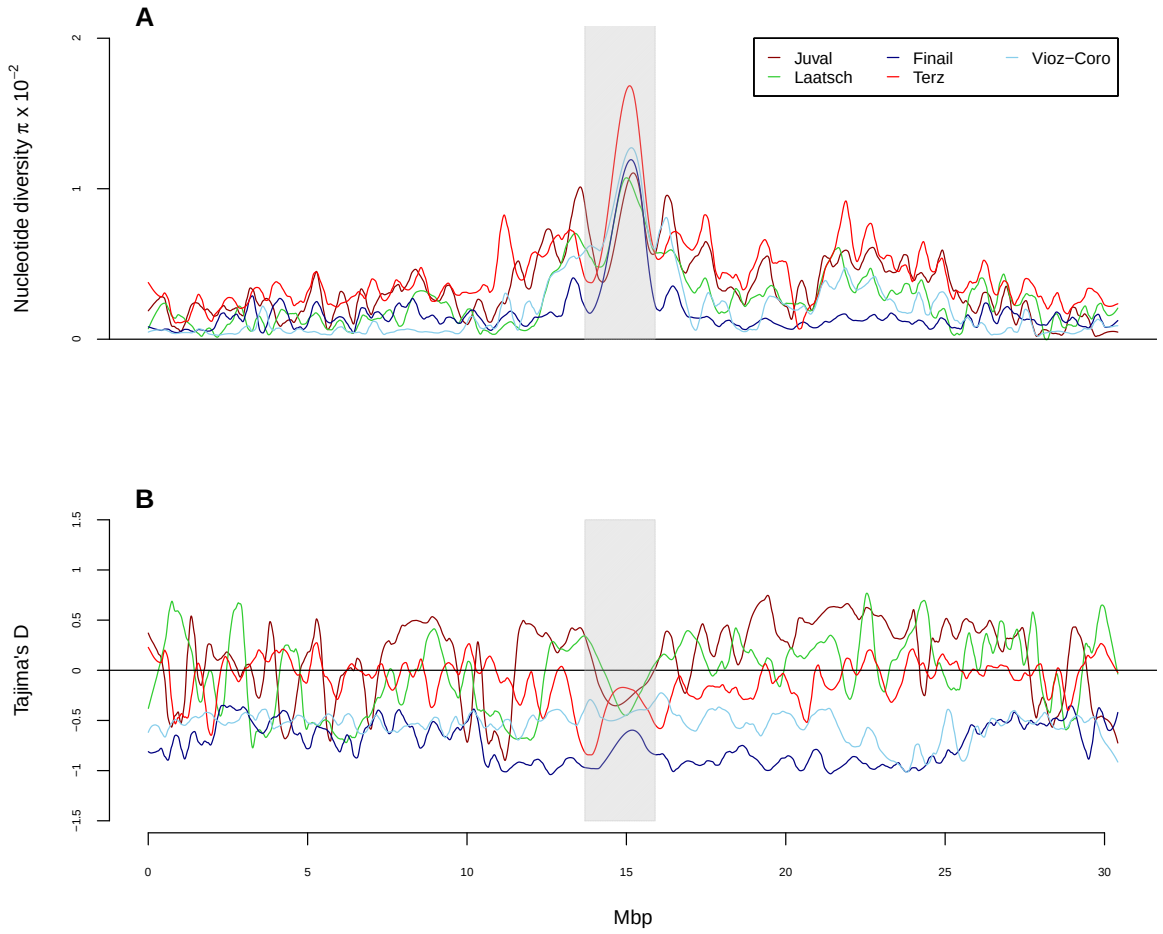

Figure 4: Sliding window analysis of (A) nucleotide diversity, π and (B) Tajima's D for chromosome 1 in all five populations.

putative functional effect (annotated as non-synonymous coding, stop lost, stop gained, start lost) segregate at much lower allele frequencies than other SNP types (Wilcoxon tests, $p < 10^{-7}$ for all classes when compared to synonymous SNPs in each population). Allele frequencies of deleterious nsSNPs (as classified by MAPP) were lower than of tolerated nsSNPs in all populations (Wilcoxon tests, $p < 10^{-5}$), indicating that purifying selection acted on this class of alleles. The ratio of nsSNPs to sSNPs, however, did not differ between high and low altitude populations (Fisher's exact test, $p > 0.05$) suggesting that the strength of selection against deleterious polymorphisms was similar in high and low altitude populations.

The pairwise correlations of allele frequencies derived from the covariance matrix estimated by Bayenv2.0 and pairwise F_{ST} values (Table S2) revealed that both matrices were correlated (Mantel-test, $r = -0.63$, $p = 0.035$). Pairwise F_{ST} values were similar between all pairs of populations, although the maximal distance was observed between Vioz/Coro and Terz populations, which belong to the same drainage system. This pair also showed the lowest correlation of allele frequencies as estimated by Bayenv2.0. Generally, the correlation estimates gave a higher variation than the F_{ST} values across population pairs (Table S2), because Bayenv2.0 accounts

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for different sequence coverage in the correlation matrix.

Neighbour-Joining trees generated from the two distance matrices indicate a closer relationship between the two high altitude populations than to the low altitude population from the same drainage system (Figure 5). The F_{ST}-based tree is almost star-like with short internal branches, whereas the tree derived from the covariance matrix shows longer internal branches and a stronger clustering of population pairs. We observed essentially the same clustering with a TreeMix analysis based on 1.07 million SNPs with a minimum coverage of 20 per population (Figure 5C). Bootstrapping results of the TreeMix analysis show a 77% confidence for the Juval and Terz population group and a migration event with 73% confidence from the Finail to the Juval population. The four-population test (Reich *et al*, 2009) rejected all possible tree topologies (all $|z| > 3$, which corresponds to $p < 0.001$), but the lowest $|z|$ was observed for the tree that groups the two high altitude populations together (Table 2). The positive test statistic for this tree suggests some admixture in at least one drainage system, most likely between Finail and Juval, consistent with the TreeMix results.

Table 2: Four-population test of two high and two low-altitude populations. It tests the probability that the relationship of four populations can be explained by one of three possible tree topologies without migration. The test was conducted with 1.95 Million SNPs with a minimum coverage of 20 in each population.

Population pairs	$f4$ statistic	Std. error	z-score
(Juval,Finail) vs. (Terz,Vioz/Coro)	0.0160	0.0019	8.28
(Juval,Terz) vs. (Finail,Vioz/Coro)	0.0032	0.0010	3.10
(Juval,Vioz/Coro) vs. (Finail,Terz)	-0.0128	0.0020	-6.38

Modelling the demographic history with ABC The tree-based analyses suggested a common origin of the two high altitude populations. We therefore considered each pair of low and high altitude populations from the same drainage system as an independent replicate of the same demographic process to estimate divergence time between high and low altitude populations. In both analyses the exponential growth model with bidirectional migration gave posterior probabilities of almost 100% (Table S4) with very similar parameter estimates. The effective population size of the high altitude populations was smaller than of the low altitude populations after the population split (Table 3), with up to a three-fold difference in population size. The 90% HPDI of the split time in the models ranged from 14,800 to 260,000 years before present, and the interval was almost identical between the two population pairs (Table 3).

Highly differentiated SNPs between high and low altitude populations To identify candidate loci for altitude specific adaptation, we correlated allele frequencies of populations with altitude using Bayenv2.0. The Z (Figure S5) statistic (Günther & Coop, 2013) was calculated for 400,231 SNPs in total after quality control (see methods section). A total of 25 SNPs showed the highest score possible for Z corresponding to a strong support for a non-zero correlation (i.e. $Z = 0.5$; Table S6). One non-synonymous SNP among them was found in the gene

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

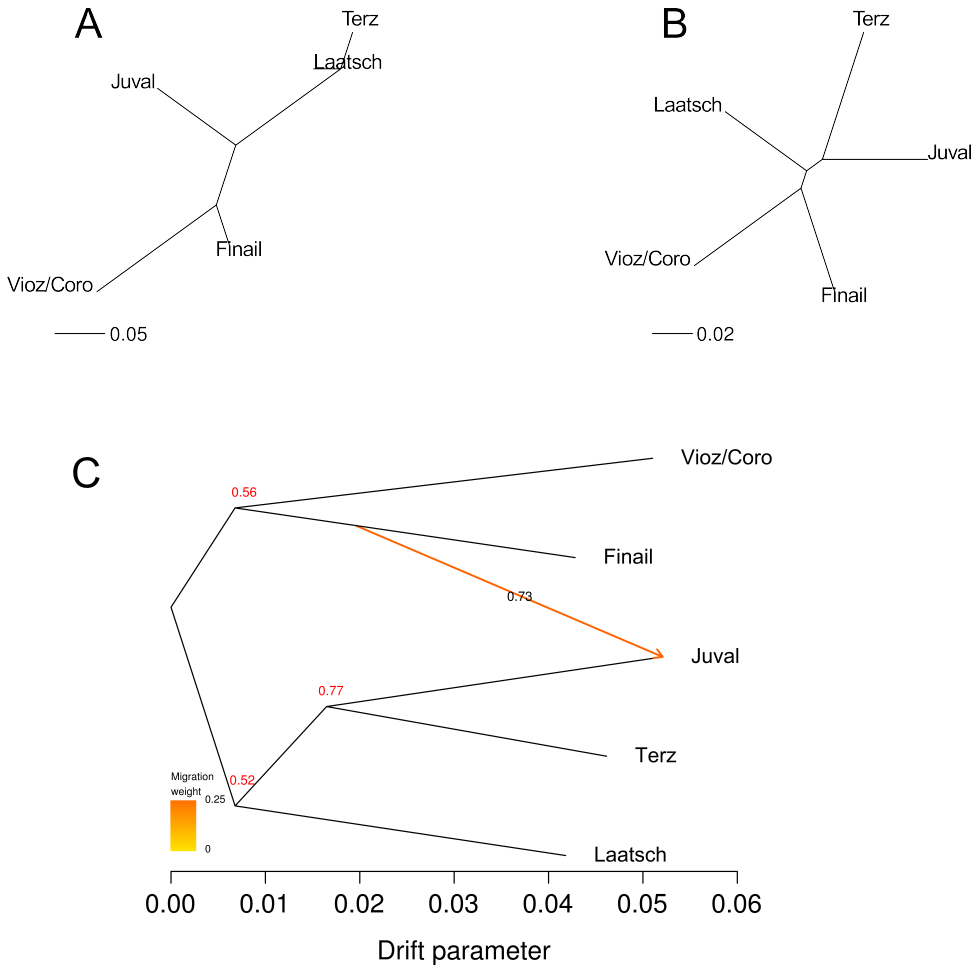

Figure 5: Relationship between populations based on measures of pairwise genetic differentiation and on allele frequencies. Branch lengths reflect the extent of genetic differentiation. (A) NJ tree of the correlation matrix obtained with Bayenv2.0. (B) NJ tree of the matrix of pairwise F_{ST} values. (C) Tree inferred with TreeMix of the five alpine populations with migration events based on 1.07 Million SNPs with a coverage of > 20. The migration event is coloured according to its weight which represents the percentage of alleles originating from the source. Numbers at branching points and on the migration arrow represent bootstrapping results based on 1,000 runs.

AT5G17860, which codes for the Calcium exchanger 7 (CAX7) protein located on chromosome 5 (Figure 6A, B). A region about 5 kb upstream of this gene showed no sequence coverage in both high altitude populations and the Laatsch population (Figure 6B, C). This pattern indicates that the Col-0 reference haplotype is not present in these populations, either because of a deletion or the presence of a highly divergent haplotype of a length of about 25 kb. The whole region was flanked by SNPs with high Z values and strong differentiation between high and low altitudes (Figure 6A). To test whether this pattern resulted from insufficient local sequence coverage, we analysed all 33,268 TAIR10 genes with sequence data in at least one population. Only 15 genes showed a similar pattern, which we defined as a per base coverage below the 5th percentile in high altitude and above the 5th percentile in low altitude populations. This

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 3: Posterior probability of the parameters of the best-fitting demographic model (exponential growth, with migration). The probability for each parameter was inferred by ABC.

	Juval - Finail		Terz - Vioz/Coro	
Parameter	Mode	90% HPDI	Mode	90% HPDI
	Exponential growth, with migration		*Exponential growth, with migration*	
$N_{\text{low alt.}}$	195,457	168,005 - 199,983	196,573	176,779 - 199,988
$N_{\text{high alt.}}$	71,221	19,460 - 166,695	77,039	18,275 - 169,580
T_{split}	70,650	18,160 - 261,636	77,491	14,892 - 260,579

low probability to find similar presence-absence variation in our sequence data supports a deletion in the high altitude populations. Of these 15 genes, 12 genes were annotated as unknown protein, protein of unknown function, pseudogene or transposon; one gene encodes a ribosomal protein, one gene a cytochrome P450 family protein and the last one is an NBS-LRR receptor (AT5G48770).

A second candidate gene encodes the Eps15 homology domain protein AtEHD1 (AT3G20290) where a single SNP located in an intron shows a high Z-score ($Z = 0.5$) but is flanked in a region of at least 50 kb by additional SNPs highly associated with altitudes. This region includes two additional genes downstream of the *AtEHD1* gene (Figure S7). In general the proportion of fixed alleles in the high-altitude populations is much larger than in the low-altitude populations. Here, only a single SNP shows the highest correlation with altitude, however, across a large region the two high-altitude populations were strongly differentiated from the other two populations.

A third example of an outlier region is located on chromosome 3 around a highly differentiated SNP in the gene AT3G07130 which encodes the purple acid phosphatase 15 (Figure S8). In this region, only the Vioz/Coro population at the highest altitude has a large number of fixed alleles and is strongly differentiated from the other four populations, which is consistent with a footprint of selective fixation of alleles in this population.

Enrichment analysis of highly differentiated genes To test whether functional groups of genes show a higher level of differentiation, we considered the top 1% of SNPs as highly differentiated SNPs and conducted an enrichment analysis with Gowinda. This analysis did not uncover enriched gene sets after correction for multiple testing with the exception of gene sets obtained from a literature collection (Lai *et al*, 2012) (Table S11). Nevertheless, categories with nominal p-values ≤ 0.05 included annotations expected to be involved in adaptation to different altitudes (Tables S7 and S8). For example, biological processes like response to light stimulus, heat acclimation, defence signalling pathway, flavonol biosynthetic process and biochemical processes like superoxide radical degradation, glucosinolate biosynthesis, flavone and flavonol biosynthesis and anthocyanin biosynthesis are some of the processes which are likely involved in adaptation to altitude. We also annotated the functional effects of highly differentiated SNPs (Figure S6). Genes highly differentiated between high and low populations were enriched for

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

genic SNPs because most SNPs were located in introns and UTRs, but not in coding regions. If genic but non-coding SNPs were involved in adaptive differentiation, selection should have preferentially acted on regulatory processes rather than on protein function. Such a hypothesis was supported by an enrichment analysis of highly differentiated SNPs in known predicted transcription factor binding (TFB) sites. According to the AGRIS database (Yilmaz *et al*, 2011) 3.3% of all SNPs were located in TFB sites, but highly differentiated SNPs were enriched in TFB sites although effect sizes were small (Bayenv2.0 candidates: 3.75%; Fisher's exact test, $p = 0.03711$).

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

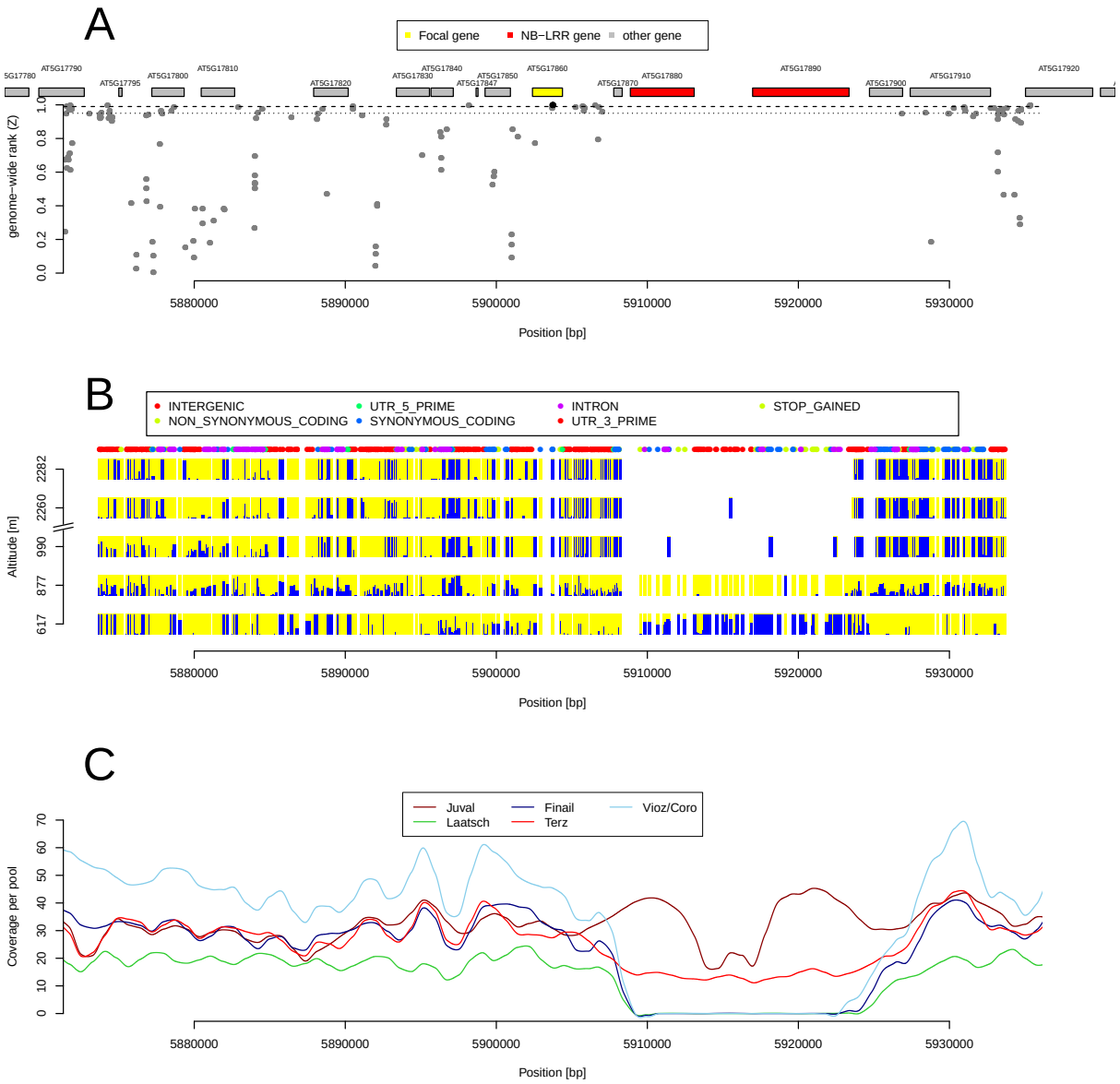

Figure 6: Example of a candidate gene (AT5G17860, encoding for *CAX7*) in which allele frequences were highly differentiated between high and low altitude populations. (A) Plot of the genome-wide relative rank of the Z statistics. High values indicate a strong differentiation. (B) Plot of the relative allele frequencies of the two alleles segregating at a polymorphic site. The height of the bar is proportional to the allele frequency. (C) Sequence coverage of the genomic region in the five different populations.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

Presence of phenotypic variation in low and high altitude populations We identified phenotypic variation in two ecologically relevant traits in the five populations but their spatial distribution did not show a strong correlation with altitude. Due to the large difference in altitude between populations (>1,000 m) and a decrease of average temperature by 0.6 °C per 100 m altitude (Barry, 1992) we expected strong selection for freezing tolerance at high altitude. Although we noted a significant variation in freezing tolerance, populations with similar frost resistance levels did not originate from the same altitude or water drainage (river) system. A possible explanation is that selection for cold tolerance is mainly influenced by microclimatic conditions which may be strongly influenced by local topology. The high altitude population Vioz/Coro is found on South-exposed rocky outcrops at protected sites underneath rock overhangs. Temperature logger data from the two high altitude sites suggest that the Vioz/Coro site is less prone to freezing during winter (Figure SS2). In addition, high alpine *A. thaliana* populations may switch to a summer annual life habit allowing the plants to avoid the period with high risk of freezing damage (Luo *et al*, 2014). Our observations of plantlets at the different localities suggest that the Vioz/Coro population consists of summer annuals, while at the other four sites plantlets germinate in the fall and persist throughout the winter.

The spatial pattern of response to light and UV-B stress differed from frost tolerance. The population at the highest altitude (Vioz/Coro) showed a lack of anthocyanin accumulation under light stress. Anthocyanins protect against cell damage by light-induced photooxidation (Chalker-Scott, 1999; Powles, 1984) and the plants from the Vioz/Coro population exhibited tissue bleaching in old roset leafs under light stress and were more sensitive to extended elevated UV-B levels. Flavonoid-deficient mutants of *A. thaliana* show a similar phenotype (Li *et al*, 1993). Solar radiation at open sky and fluctuations in radiation intensity increase with higher altitude (Blumthaler *et al*, 1997; Körner, 2007) and high altitude populations need to adapt to more intense UV-B and light conditions. For example, the Shahkdara accession of *A. thaliana* was collected at 3,063 m a.s.l. and exhibited a constitutive UV-B protection (Biswas & Jansen, 2012). We therefore suggest that the deficiency in anthocyanin accumulation in the Vioz/Coro population is not an adaptation to high altitudes but a loss-of-function phenotype resulting from genetic drift at the range margins (Eckert *et al*, 2008).

In summary, the spatial patterns of variation in the two traits do not support a simple pattern of adaptation to altitude because we expected both high altitude populations to show a similar phenotype and the demographic analysis suggested that there was sufficient time available for adaptation to high altitude. Possible explanations are that a difference of 1,700 m altitude between sampling locations does not represent a strong selection gradient or that the microenvironment or random drift and metapopulation dynamics of small demes confounds clinal trait variation.

Patterns of genetic diversity and population differentiation In *A. thaliana*, levels of genetic diversity differ strongly among populations at small and large geographical scales

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Bomblies *et al*, 2010; Cao *et al*, 2011). A low genetic diversity limits the potential to respond to selection and may lead to the accumulation of deleterious alleles (Alleaume-Benharira *et al*, 2006). In our sample, both high altitude populations exhibited reduced genetic diversity and genome-wide negative Tajima's D values that indicate a deviation from a standard neutral model. This pattern may result from bottlenecks with subsequent population growth (Simonsen *et al*, 1995) as expected after the colonization of new habitats like high altitude sites at the outer vertical limit of a species distribution. However, we did not observe an increase of deleterious amino acid polymorphisms in high altitude populations that would accompany such a bottleneck due to a reduction of purifying selection (Günther & Schmid, 2010).

Although diversity estimates differed significantly among the five populations, we found little evidence for a strong population structure. The average pairwise F_{ST} value of 0.128 was three times smaller than in local populations near Tübingen, Germany ($F_{ST} = 0.52$; Bomblies *et al*, 2010) and smaller than the smallest observed pairwise F_{ST} in a sample of Swiss alpine *A. thaliana* populations that were based on 24 SSR loci (Luo *et al*, 2014). In addition, pairwise F_{ST} values were very similar for all pairs of populations, consistent with a low number of unique SNPs per accession and a high level of LD in a different sample of South Tyrolian accessions (Cao *et al*, 2011). The demographic history of numerous alpine plant species was influenced by glaciation which caused repeated cycles of retreats into refugia and colonization of newly available habitats in interglacial periods (Schönswetter *et al*, 2005). Refugial populations in alpine valleys or local refugia on mountain tops (nunataks) above glaciers may have contributed to a strong genetic differentiation between subpopulations. A low differentiation and a star-like relationship with gene flow between populations (e.g., the low-altitude Juval and the high-altitude Finail populations; Figure 5) of our five populations is not consistent with such a scenario and instead suggest that these sites were colonized recently from refugia at the edge or outside the alps. On the other hand, the two high altitude populations were more closely related to each other than to the respectice low-altitude populations from the same river drainage. This may reflect an origin from the same (mountaintop) refugium or a recent dispersal by animals. Generally, a rapid postglacial recolonization of alpine habitats by *A. thaliana* is possible because of the mainly self-fertilizing mode of reproduction and dispersal by animals like mountain goats. The high altitude populations were found only at highly disturbed resting sites of mountain goats.

To investigate the age of the split of low and high-altitude populations, we estimated demographic parameters for different demographic models with ABC (Beaumont *et al*, 2002). Both pairs of low and high altitude population supported a model with a smaller effective population size in high altitude populations caused by a bottleneck and subsequent population growth. The 90% HDPI intervals for population split times ranged between 14,000 and 260,000 years before present. This includes time before and after the last glacial maximum (LGM) as deglaciation on the Northern Hemisphere began about 19-20 thousand years ago (Clark *et al*, 2009). Consequently, our ABC analysis did not allow to answer the question whether the split between low- and high-altitude populations occurred before or after the LGM, though the probability is

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

higher that it occurred before. The wide HDPI 90 interval may result from a limitation of pool sequencing because summary statistics used for ABC were limited to measures of gene frequencies and did not include haplotype-based statistics, which provide additional power for inferring population split times and admixture levels by quantifying the length of shared fragments (Sved *et al*, 2008; Sankararaman *et al*, 2012; Hellenthal *et al*, 2014).

Patterns of genetic variation and local selective sweeps In *A.thaliana* only few species-wide selective sweeps were found so far (Clark *et al*, 2007; Childs *et al*, 2010; Cao *et al*, 2011; Horton *et al*, 2012), and the overall frequency of adaptive evolution is very low or absent based on genome-wide polymorphism data (Gossmann *et al*, 2010; Slotte *et al*, 2011) although evidence for local adaptation was found in several studies (Fournier-Level *et al*, 2011; Hancock *et al*, 2011; Agren & Schemske, 2012; Méndez-Vigo *et al*, 2011; Long *et al*, 2013; Huber *et al*, 2014). Therefore, searching for signatures of selection at smaller geographic scales along environmental gradients may identify phenotypic and genomic footprints of adaptive evolution. We identified candidate genes for high-altitude adaptation with an outlier approach. It should be noted that this is no formal test of selection and candidate genes require further validation (Pavlidis *et al*, 2012).

Patterns of genetic variation at some candidate genes suggested that targets of selection were difficult to elucidate because large genomic regions (>50k) around significantly associated SNPs involving multiple adjacent genes showed a population-specific sweep pattern (Figures 6, S7 and S8). The comparison of the three genomic regions showed that despite an overall low level of genetic diversity in the alpine populations, complex patterns of genetic diversity that involved presence/absence variants (PAVs) and regions of high nucleotide diversity were observed. For example, in our sample, a single SNP was significantly differentiated between low and high altitude populations and some neighboring SNPs also showed a clinal pattern (Figures 6). It was previously identified in *Arabidopsis lyrata* as a candidate selection target for adaptation to soils with a low Ca:Mg-ratio (Turner *et al*, 2008) and *A. thaliana* is generally found on siliceous rock with low Ca content. However, in the direct neighbourhood, although pool sequencing did not allow the analysis of haplotypes, patterns of sequence coverage strongly suggested that a PAV comprising two well characterized NBS-LRR genes, *CSA1* and *CHS3*, may have contributed to local adaptation since the deletion was fixed in the high altitude and Laatsch populations, whereas it segregated in the low-altitude populations (Figure 6). In addition both genes showed a PAV across the whole species because *CSA1* was not covered by sequence reads in 11 and *CHS3* in 9 out of 18 high coverage reference genomes of a diverse set of accessions (Gan *et al*, 2011). PAVs are abundant across the genome of *A.thaliana* (Tan *et al*, 2012), particularly in resistance genes that include NBS-LRR genes (Bergelson *et al*, 2001; Shen *et al*, 2006; Guo *et al*, 2011; Karasov *et al*, 2014). It remains open whether selection or genetic drift contributed to the fixation at these genes. They are involved in pathogen response, but *CSA1* additionally plays a role in shade avoidance and response to red light (Faigón-Soverna *et al*, 2006) and knockouts of *CHS3* cause chilling-sensitive plants (Schneider *et al*, 1995). A comparison of sequence coverage pattern against all genes annotated in TAIR10 showed, however, that a similar deletion pattern

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

was observed in only 15 genes that included other NBS-LRR genes or genes of unknown function.

The pattern of diversity at the *AtEHD1* gene (AT3G20290) may represent a population-specific selective sweep because of numerous SNPs with a high Z-score in this region. The same alleles are fixed in the two high-altitude populations and the diversity is strongly reduced over a range of about 50 kb (Figure S7). In contrast, the low altitude populations appeared to be highly polymorphic in this region. The identification of putative selection targets requires further study, but it is noteworthy that *AtEHD1* increases tolerance to salt stress (Bar *et al*, 2013).

We observed another pattern of genetic diversity in the gene encoding the purple acid phosphatase 15 (AT3G07130) where only the highest population was differentiated (Figure S8). This gene may affect phosphorus efficiency by remobilizing inorganic phosphate from organic phosphate sources (Wang *et al*, 2009) and is a plausible selection target because phosphorus was highest in the highest population (Table S5). Taken together, diversity patterns in several genomic regions suggest that adaptation to soil conditions may have played a role in the history of these populations (Baxter *et al*, 2010).

Changes in light conditions and temperature differences are among the expected environmental differences between altitudes. We did not identify genes known to be involved in frost resistance in the GO enrichment analysis (Table S7), which corresponds to the patterns of phenotypic differentiation. However, genes associated with heat acclimation were among the identified candidates, which may relate to the observation that high temperatures in late spring terminate the reproductive cycle in all low altitude populations by causing plant death. In contrast, mild temperatures at high altitudes allow reproduction throughout the summer suggesting selection for heat acclimation may be relaxed (personal observation: C. Lampei). Furthermore, in several genes involved in shade avoidance and response to light stimulus as well as mutations in the anthocyanin (Table S8), flavonol and phenylpropanoid biosynthesis pathways (Table S7) were correlated with altitude. This pattern is consistent with phenotypic differentiation, although the anthocyanin deficiency may constitute a loss of function mutation in the highest population, indicating that genetic differentiation at some genes may not result from past selection.

In a recent genome-wide scan in *A. halleri*, genes associated with red or far red light response were differentiated with altitude in two independent altitude gradients in Japan (Kubota *et al*, 2015). Further, genes involved in defence response were also enriched in an Alpine altitude gradient study in the same plant (Fischer *et al*, 2013). Notably, the enrichment analysis did not identify any soil related categories among the nominally significant annotations, which is in strong contrast to the annotation of the top 25 highly differentiated genes. This corresponds to the observation by Fischer *et al* (2013) that the GO enrichment analysis did not match with outlier correlations with environmental variables.

Methods to identify selection along altitudinal gradients Pool sequencing is a cost-efficient method to characterize genetic variation using allele frequency estimates. Reliable estimates of population-specific allele frequencies require a sufficient sample size and sequence coverage (Schlötterer *et al*, 2014). Our sequenced pools fulfill recommendations for pool sequenc-

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ing and we expect that the observed patterns of genetic diversity resemble genomic differences and not artefacts of the method. Both our phenotypic and genomic analysis identified substantial variation along altitudinal gradients. Studies of local adaptation need to consider the effect of sampling design on the probability to identify false positives (Meirmans, 2015). Simulations suggest that studies of local adaptation should be based on several pairs of populations that differ in the environmental variable of interest (Lotterhos & Whitlock, 2015). We included two population pairs to have to samples of putative adaptive processes and demographic history. The closer relationship of the high-altitude populations, however, suggest that genetic variation shares a joint history and are not entirely independent. For this reason, the power to identify targets of altitudinal adaptation increases with more population pairs, if local populations are small and isolated and therefore prone to genetic drift. Future studies of altitudinal adaptation in *A. thaliana* should therefore be based on a denser sampling of populations and include additional phenotypic traits (Montesinos-Navarro *et al*, 2011; Luo *et al*, 2014). Given the patchy and highly disjunct distribution of *A. thaliana* in the high Alps, this might require systematic sampling over a larger geographic area to disentangle demography from natural selection (Meirmans *et al*, 2011).

Conclusion The phenotypic variation in response to two environmental variables, which predictably change with altitude, was not consistent with selection for high-altitude adaptation in our *A. thaliana* populations from the North Italian Alps. Microevolutionary responses to local topology at the population sites, founder effects and small population sizes at high altitude may overrule mean altitudinal change in environmental variables. In contrast, several genomic regions showed patterns of genetic variation consistent with population-specific selection-driven sweeps, that were associated with observed environmental differences (e.g. soil parameters). Furthermore, the detection of outlier genes suggested some plausible targets positive selection. These candidate genes can be tested at a functional level in future studies to test their potential role in adaptation to high altitude. Patterns of genetic diversity indicated a closer relationship of the high altitude populations that may reflect a common demographic history during glacial-interglacial cycles.

Acknowledgements

We thank Elisabeth Kokai-Kota for lab assistance, Hinrich Bremer for access to a conductivity meter, Dr. Nikolaus Merkt for access to a UV-light meter and Fabian Freund for statistical discussions. This work was funded by the ESF EUROCORES project EpiCol (SCHM 1354/4-1), the German Federal Ministry for Education and Research (BMBF) within the AgroClustEr Synbreed—Synergistic plant and animal breeding (FKZ0315528D), and BMBF Plant2030 project RYE-SELECT (FKZ 0315946E).

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Agren J, Schemske DW (2012) Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. *The New Phytologist*, **194**, 1112–22.

Al-Shehbaz Ia, O'Kane SL (2002) Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae). *The Arabidopsis book / American Society of Plant Biologists*, **1**, e0001.

Alleaume-Benharira M, Pen IR, Ronce O (2006) Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *Journal of Evolutionary Biology*, **19**, 203–15.

Altschul SF, Madden TL, Schaffer AA, *et al* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Ashburner M, Ball CA, Blake JA, *et al* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, **25**, 25–29.

Bar M, Leibman M, Schuster S, Pitzhadza H, Avni A (2013) EHD1 functions in endosomal recycling and confers salt tolerance. *PloS ONE*, **8**, e54533.

Barry RG (1992) *Mountain Weather and Climate*. Routledge, London, 2nd edn..

Baxter I, Brazelton JN, Yu D, *et al* (2010) A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1; 1. *PLoS Genetics*, **6**, e1001193.

Beaumont Ma, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–35.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, **57**, 289–300.

Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant R-genes. *Science*, **292**, 2281–5.

Bergelson J, Roux F (2010) Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nature Reviews Genetics*, **11**, 867–879.

Biswas DK, Jansen MAK (2012) Natural variation in UV-B protection amongst *Arabidopsis thaliana* accessions. *Emirates Journal of Food and Agriculture*, **24**, 621–631.

Blumthaler M, Ambach W, Ellinger R (1997) Increase in solar UV radiation with altitude. *Journal of Photochemistry and Photobiology B: Biology*, **39**, 130–134.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Boitard S, Schlötterer C, Nolte V, Pandey RV, Futschik A (2012) Detecting selective sweeps from pooled next-generation sequencing samples. *Molecular Biology and Evolution*, **29**, 2177–86.

Bomblies K, Yant L, Laitinen Ra, *et al* (2010) Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genetics*, **6**, e1000890.

Cao J, Schneeberger K, Ossowski S, *et al* (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics*, **43**, 956–963.

Chalker-Scott L (1999) Environmental significance of anthocyanins in plant stress responses. *Photochemistry and Photobiology*, **70**, 1–9.

Childs LH, Witucka-Wall H, Günther T, *et al* (2010) Single feature polymorphism (SFP)-based selective sweep identification and association mapping of growth-related metabolic traits in *Arabidopsis thaliana*. *BMC Genomics*, **11**, 188.

Cingolani P, Platts A, Wang LL, *et al* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, **6**, 80–92.

Clark PU, Dyke AS, Shakun JD, *et al* (2009) The Last Glacial Maximum. *Science*, **325**, 710–714.

Clark RM, Schweikert G, Toomajian C, *et al* (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science*, **317**, 338–342.

Cooper L, Walls RL, Elser J, *et al* (2013) The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant & Cell Physiology*, **54**, e1.

Cox MP, Peterson Da, Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*, **11**, 485.

Csilléry K, François O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, **3**, 475–479.

DevelopmentCoreTeam (2014) *R: A Language and Environment for Statistical Computing*. Vienna.

Eckert C, Samis K, Lougheed S (2008) Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170–1188.

Etterson JR, Shaw RG (2001) Constraint to adaptive evolution in response to global warming. *Science*, **294**, 151–154.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

680 Fabian DK, Kapun M, Nolte V, *et al* (2012) Genome-wide patterns of latitudinal differentiation
681 among populations of *Drosophila melanogaster* from North America. *Molecular Ecology*, **21**,
682 4748–69.

683 Faigón-Soverna A, Harmon FG, Storani L, *et al* (2006) A constitutive shade-avoidance mutant
684 implicates TIR-NBS-LRR proteins in *Arabidopsis* photomorphogenic development. *The Plant*
685 *Cell*, **18**, 2919–28.

686 Fischer MC, Rellstab C, Tedder A, *et al* (2013) Population genomic footprints of selection
687 and associations with climate in natural populations of *Arabidopsis halleri* from the Alps.
688 *Molecular Ecology*, pp. 5594–5607.

689 Fournier-Level a, Korte a, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of
690 local adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86–9.

691 Friedman N, Ninio M, Pe'er I, Pupko T (2002) A structural EM algorithm for phylogenetic
692 inference. *Journal of Computational Biology*, **9**, 331–353.

693 Futschik A, Schlötterer C (2010) The next generation of molecular markers from massively
694 parallel sequencing of pooled DNA samples. *Genetics*, **186**, 207–18.

695 Gan X, Stegle O, Behr J, *et al* (2011) Multiple reference genomes and transcriptomes for
696 *Arabidopsis thaliana*. *Nature*, **477**, 419–23.

697 Gossmann TI, Song BH, Windsor AJ, *et al* (2010) Genome wide analyses reveal little evidence
698 for adaptive evolution in many plant species. *Molecular Biology and Evolution*, **27**, 1822–32.

699 Günther T, Coop G (2013) Robust Identification of Local Adaptation from Allele Frequencies.
700 *Genetics*, **195**, 205–220.

701 Günther T, Schmid KJ (2010) Deleterious amino acid polymorphisms in *Arabidopsis thaliana*
702 and rice. *Theoretical and Applied Genetics*, **121**, 157–68.

703 Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D (2011) Genome-wide comparison
704 of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. *Plant Physiology*,
705 **157**, 757–69.

706 Hall B (2013) LaplacesDemon: An R Package for Bayesian Inference, version 13.11.17.

707 Hancock AM, Brachi B, Faure N, *et al* (2011) Adaptation to climate across the *Arabidopsis*
708 *thaliana* genome. *Science*, **334**, 83–6.

709 He Z, Zhai W, Wen H, *et al* (2011) Two evolutionary histories in the genome of rice: the roles
710 of domestication genes. *PLoS Genetics*, **7**, e1002100.

711 Hellenthal G, Busby GBJ, Band G, *et al* (2014) A genetic atlas of human admixture history.
712 *Science*, **343**, 747–751.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Horton MW, Hancock AM, Huang YS, *et al* (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nature Genetics*, **44**, 212–6.

Hu TT, Pattyn P, Bakker EG, *et al* (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics*, **43**, 476–481.

Huber CD, Nordborg M, Hermisson J, Hellmann I (2014) Keeping it local: Evidence for positive selection in Swedish *Arabidopsis thaliana*. *Molecular Biology and Evolution*, **31**, 3026–3039.

Huxley JS (1938) Clines: an auxiliary method in taxonomy. *Nature*, **142**, 219–220.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, **40**, gkr988.

Karasov TL, Horton MW, Bergelson J (2014) Genomic variability as a driver of plant-pathogen coevolution? *Current Opinion in Plant Biology*, **18**, 24–30.

Kim S, Plagnol V, Hu TT, *et al* (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics*, **39**, 1151–1155.

Kofler R, Betancourt AJ, Schlötterer C (2012) Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genetics*, **8**, –1002487.

Kofler R, Orozco-terWengel P, De Maio N, *et al* (2011a) PoPoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE*, **6**, e15925.

Kofler R, Pandey RV, Schlötterer C (2011b) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, **27**, 3435–6.

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ (2011) Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics*, **187**, 245–60.

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology*, **55**, 141–72.

Körner C (2007) The use of 'altitude' in ecological research. *Trends in Ecology & Evolution*, **22**, 569–74.

Kubota S, Iwasaki T, Hanada K, *et al* (2015) A Genome Scan for Genes Underlying Microgeographic-Scale Local Adaptation in a Wild Arabidopsis Species. *PLOS Genetics*, **11**, e1005361.

Lai L, Liberzon A, Hennessey J, *et al* (2012) AraPath: a knowledgebase for pathway analysis in *Arabidopsis*. *Bioinformatics*, **28**, 2291–2.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

747 Lamesch P, Berardini TZ, Li D, *et al* (2012) The *Arabidopsis* Information Resource (TAIR):
748 improved gene annotation and new tools. *Nucleic Acids Research*, **40**, D1202–D1210.

749 Lamichhaney S, Martinez Barrio A, Rafati N, *et al* (2012) Population-scale sequencing reveals
750 genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National*
751 *Academy of Sciences of the United States of America*, **109**, 19345–50.

752 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform.
753 *Bioinformatics*, **25**, 1754–60.

754 Li H, Handsaker B, Wysoker A, *et al* (2009) The Sequence Alignment/Map format and SAM-
755 tools. *Bioinformatics*, **25**, 2078–9.

756 Li J, Ou-Lee TM, Raba R, Amundson RG, Last RL (1993) Arabidopsis Flavonoid Mutants Are
757 Hypersensitive to UV-B Irradiation. *The Plant Cell*, **5**, 171–179.

758 Long Q, Rabanal Fa, Meng D, *et al* (2013) Massive genomic variation and strong selection in
759 *Arabidopsis thaliana* lines from Sweden. *Nature Genetics*, **45**, 884–90.

760 Lotterhos K, Whitlock MC (2015) The relative power of genome scans to detect local adaptation
761 depends on sampling design and statistical method. *Molecular Ecology*, **24**, 1031–1046.

762 Luo Y, Dong X, Yu TY, *et al* (2015) A single nucleotide deletion in *GA20ox1* causes alpine
763 dwarfism in *Arabidopsis thaliana*. *Plant Physiology*, pp. pp–00005.

764 Luo Y, Widmer A, Karrenberg S (2014) The roles of genetic drift and natural selection in
765 quantitative trait divergence along an altitudinal gradient in *Arabidopsis thaliana*. *Heredity*,
766 pp. 220–228.

767 Manel S, Gugerli F, Thuiller W, *et al* (2012) Broad-scale adaptive genetic variation in alpine
768 plants is driven by temperature and precipitation. *Molecular Ecology*, **21**, 3729–38.

769 Martin M, Gavazov K, Körner C, Hättenschwiler S, Rixen C (2010) Reduced early growing
770 season freezing resistance in alpine treeline plants under elevated atmospheric CO 2. *Global*
771 *Change Biology*, **16**, 1057–1070.

772 Mayr E (1942) *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. Harvard
773 University Press, New York.

774 Meirmans PG (2015) Seven common mistakes in population genetics and how to avoid them.
775 *Molecular Ecology*.

776 Meirmans PG, Goudet J, Gaggiotti OE (2011) Ecology and life history affect different aspects
777 of the population structure of 27 high-alpine plants. *Molecular Ecology*, **20**, 3144–3155.

778 Méndez-Vigo B, Picó FX, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C (2011) Altitudinal
779 and climatic adaptation is mediated by flowering traits and FRI, FLC, and PHYC genes in
780 *Arabidopsis*. *Plant Physiology*, **157**, 1942–55.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

781 Montesinos-Navarro A, Wig J, Xavier Pico F, Tonsor SJ (2011) *Arabidopsis thaliana* populations
782 show clinal variation in a climatic gradient associated with altitude. *The New Phytologist*,
783 **189**, 282–94.

784 Mueller LA, Zhang P, Rhee SY (2003) AraCyc: a biochemical pathway database for *Arabidopsis*.
785 *Plant Physiology*, **132**, 453–60.

786 Orozco-terWengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlötterer C (2012) Adaptation of
787 *Drosophila* to a novel laboratory environment reveals temporally heterogeneous trajectories
788 of selected alleles. *Molecular Ecology*, **21**, 4931–41.

789 Ossowski S, Schneeberger K, Lucas-Lledó JI, *et al* (2010) The rate and molecular spectrum of
790 spontaneous mutations in *Arabidopsis thaliana*. *Science*, **327**, 92–94.

791 Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R
792 language. *Bioinformatics*, **20**, 289–90.

793 Park JS, Choung MG, Kim JB, *et al* (2007) Genes up-regulated during red coloration in UV-B
794 irradiated lettuce leaves. *Plant Cell Reports*, **26**, 507–16.

795 Patterson N, Moorjani P, Luo Y, *et al* (2012) Ancient admixture in human history. *Genetics*,
796 **192**, 1065–1093.

797 Pavlidis P, Jensen JD, Stephan W, Stamatakis A (2012) A critical assessment of storytelling:
798 gene ontology categories and the importance of validating genomic scans. *Molecular Biology*
799 *and Evolution*, **29**, 3237–48.

800 Pavlidis P, Laurent S, Stephan W (2010) msABC: a modification of Hudson's ms to facilitate
801 multi-locus ABC analysis. *Molecular Ecology Resources*, **10**, 723–727.

802 Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide
803 allele frequency data. *PLoS Genetics*, **8**, e1002967.

804 Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC (2014) nlme: Linear and Nonlinear Mixed
805 Effects Models. R package version 3.1-117.

806 Powles SB (1984) Photoinhibition of Photosynthesis Induced by Visible Light. *Annual Review*
807 *of Plant Physiology*, **35**, 15–44.

808 Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population
809 history. *Nature*, **461**, 489–94.

810 Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylo-
811 genetic trees. *Molecular Biology and Evolution*, **4**, 406–25.

812 Sankararaman S, Patterson N, Li H, Pääbo S, Reich D (2012) The date of interbreeding between
813 Neandertals and modern humans. *PLoS Genetics*, **8**, e1002947.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

814 Schindelin J, Arganda-Carreras I, Frise E, *et al* (2012) Fiji: an open-source platform for
815 biological-image analysis. *Nature methods*, **9**, 676–682.

816 Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals — mining
817 genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, **15**, 749–763.

818 Schneider JC, Suzanne H, Somerville CR (1995) Chilling-sensitive mutants of *Arabidopsis*. *Plant*
819 *Molecular Biology Reporter*, **13**, 11–17.

820 Schönswetter P, Stehlik I, Holderegger R, Tribsch A (2005) Molecular evidence for glacial refugia
821 of mountain plants in the European Alps. *Molecular Ecology*, **14**, 3547–55.

822 Shen J, Araki H, Chen L, Chen JQ, Tian D (2006) Unique evolutionary mechanism in R-genes
823 under the presence/absence polymorphism in *Arabidopsis thaliana*. *Genetics*, **172**, 1243–50.

824 Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for
825 DNA polymorphism data. *Genetics*, **141**, 413–29.

826 Slotte T, Bataillon T, Hansen TT, St Onge K, Wright SI, Schierup MH (2011) Genomic determi-
827 nants of protein evolution and polymorphism in *Arabidopsis*. *Genome Biology and Evolution*,
828 **3**, 1210–9.

829 Stinchcombe JR, Weinig C, Ungerer M, *et al* (2004) A latitudinal cline in flowering time in
830 *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. *Proceedings of the*
831 *National Academy of Sciences of the United States of America*, **101**, 4712–7.

832 Stone EA, Sidow A (2005) Physicochemical constraint violation by missense substitutions me-
833 diates impairment of protein function and disease severity. *Genome Research*, **15**, 978–986.

834 Sved JA, McRae AF, Visscher PM (2008) Divergence between human populations estimated
835 from linkage disequilibrium. *American Journal of Human Genetics*, **83**, 737–743.

836 Tan S, Zhong Y, Hou H, Yang S, Tian D (2012) Variation of presence/absence genes among
837 *Arabidopsis populations*. *BMC Evolutionary Biology*, **12**, 86.

838 Thiel-Egenter C, Alvarez N, Holderegger R, *et al* (2011) Break zones in the distributions of
839 alleles and species in alpine plants. *Journal of Biogeography*, **38**, 772–782.

840 Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing
841 reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics*, **42**, 260–3.

842 Turner TL, von Wettberg EJ, Nuzhdin SV (2008) Genomic analysis of differentiation between
843 soil types reveals candidate genes for local adaptation in *Arabidopsis lyrata*. *PloS ONE*, **3**,
844 e3183.

845 Wang X, Wang Y, Tian J, Lim BL, Yan X, Liao H (2009) Overexpressing AtPAP15 enhances
846 phosphorus efficiency in Soybean. *Plant Physiology*, **151**, 233–40.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

847 Wu CH, Apweiler R, Bairoch A, *et al* (2006) The Universal Protein Resource (UniProt): an
848 expanding universe of protein information. *Nucleic Acids Research*, **34**, D187–D191.

849 Yilmaz A, Mejia-Guerra MK, Kurz K, Liang X, Welch L, Grotewold E (2011) AGRIS: the
850 *Arabidopsis* Gene Regulatory Information Server, an update. *Nucleic Acids Research*, **39**,
851 D1118–D1122.

852 Zhu Y, Bergland AO, González J, Petrov DA (2012) Empirical validation of pooled whole
853 genome population re-sequencing in *Drosophila melanogaster*. *PLoS ONE*, **7**, e41901.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Data availability The raw sequence data were deposited in the Short Read Archive (XXXXX). SNP calls and other downstream sequence analyse data were deposited with with all phenotypic measurements and the R and Python scripts used for the analysis at DataDryad under accession number XXXXX.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Phenotypic and genomic differentiation along altitudinal gradients in South-Alpine populations of *Arabidopsis thaliana*

Supporting information

Torsten Günther[$,‡,§], Christian Lampei[$,‡], Ivan Barilar[$], and Karl J. Schmid[$,*]

July 22, 2015

[$] Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany
[§]Department of Evolutionary Biology, EBC, Uppsala University, Uppsala, Sweden.
[*]Corresponding author: E-mail: karl.schmid@uni-hohenheim.de.
[‡]These authors contributed equally to this work.

Supplementary Methods

Resistance to high UV-B radiation To test resistance to higher UV-B dosages we subjected 44 days old plants (10 pots and genotypes per population) which were still small due to 19 days of vernalization (4℃ and short day (8h light)) to increasing UV-B dosages of continuous UV-B light over 3 days. The UV-B levels were $0, 2.47, 6.39$ and $15.16\ kJm^{-2}day^{-1}$. After this treatment plants were left to recover for 8 days to improve the separation of dead and alive biomass. All plants were evaluated manually with their identity hidden from the investigator. Dead and living tissue was separated for each pot and oven dried for 12 h at 90 ℃ before weighting. A GLM with dead biomass odds as response variable and the fixed effects population and UV-B dosage was fitted with a quasibinomial error distribution (logit link) in R (DevelopmentCoreTeam, 2014, package:stats).

Soil properties In all sites soil samples of the top soil were collected. The soil samples were composed from about 40 individual samples per population site which were collected with a hand shovel from the uppermost 5 cm well spread across the distribution of the population. The soil was mixed, homogenized and analysed for standard parameters (Table S5).

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figures

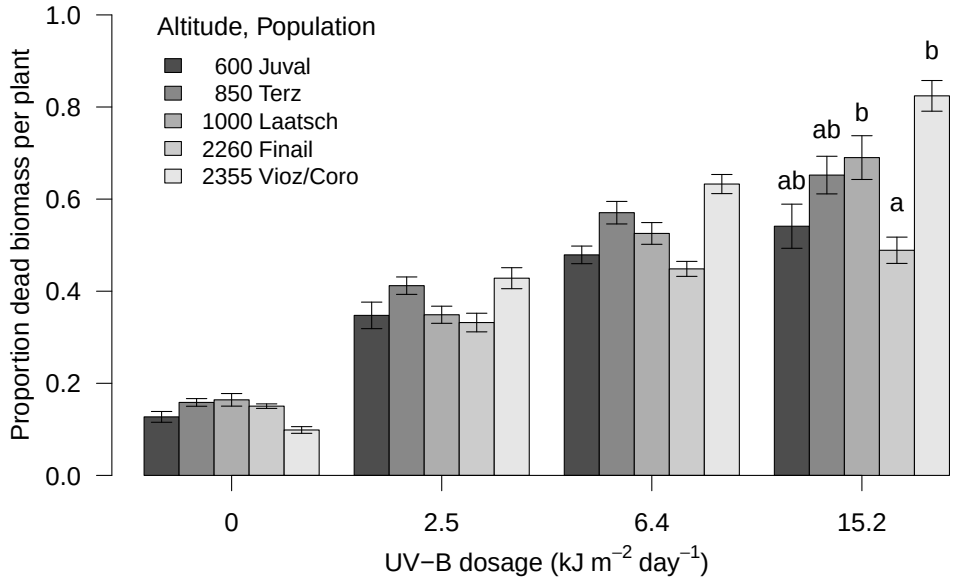

Figure S1: Proportion of dead tissue after 3 days continuous UV-B treatment with different dosages. Letters above error bars indicate homogeneous subsets after correction for false discovery rate for the highest UV-B dosage

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

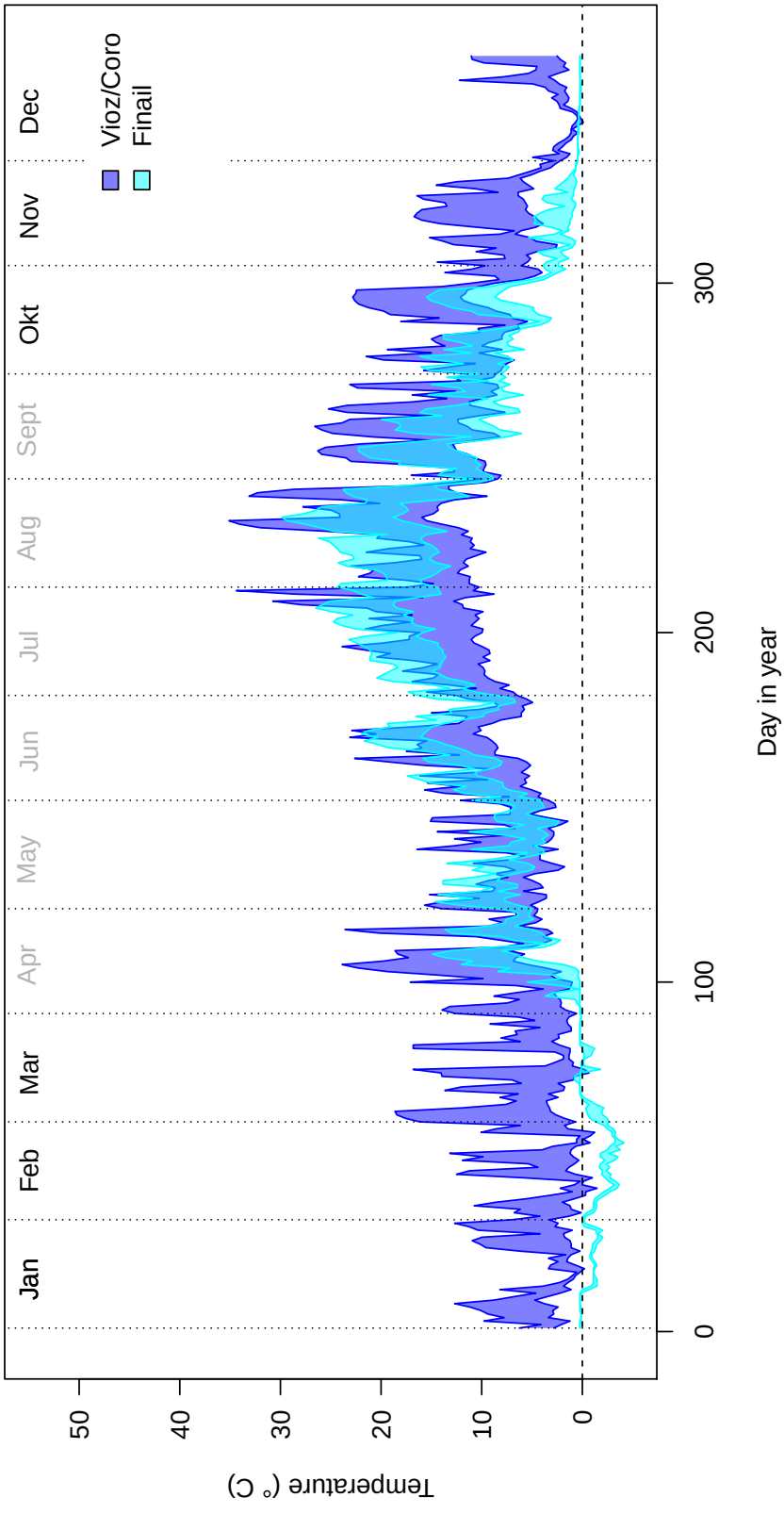

Figure S2: Daily minimum and maximum temperature for the top soil layer (< 5 cm) in the two high altitude sites

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

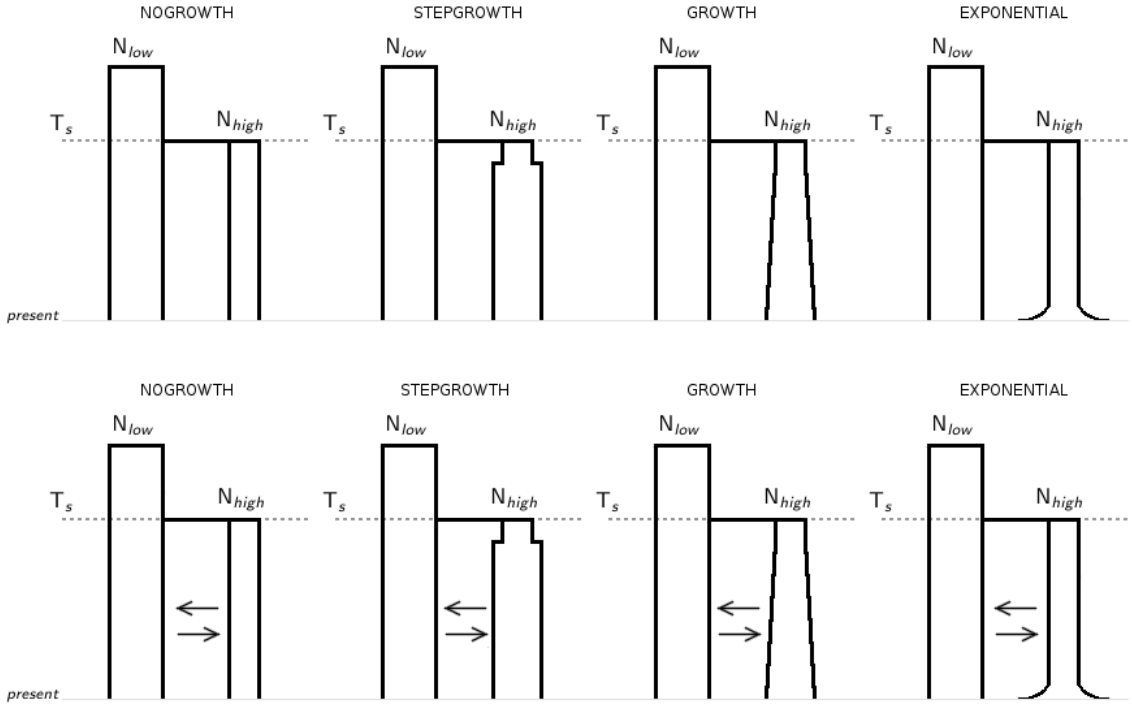

Figure S3: Four demographic models without and with migration investigated in this study. NOGROWTH - Model without the growth of the high altitude population after the split. STEPGROWTH - Model with one instantaneous size change of the high altitude population after the split. GROWTH - Model with simulated growth of the high altitude population after the split. EXPONENTIAL - Model with an exponential growth of the high altitude population sometime after the split. T_s - split time. N_{low} and N_{high} - effective population size of the low and high altitude population, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

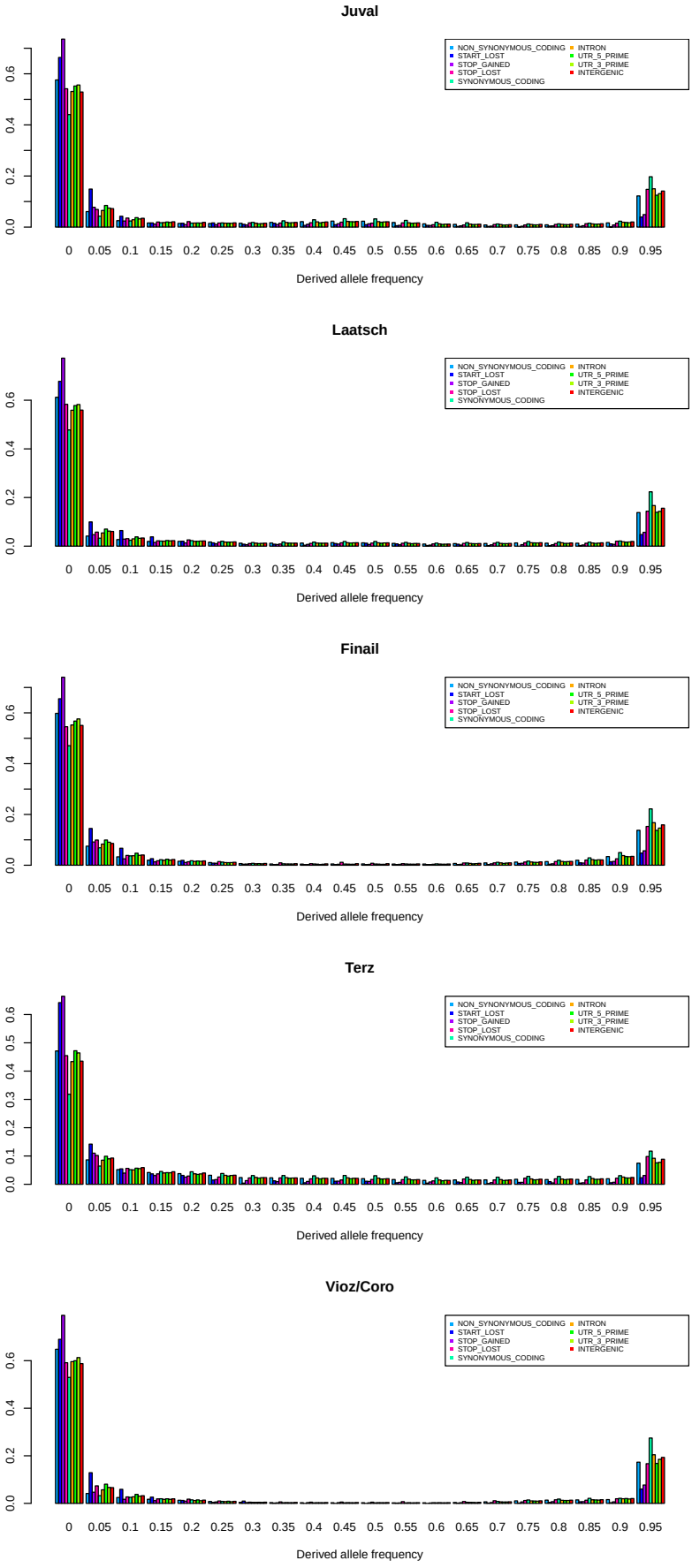

Figure S4: Derived allele frequency distributions across all populations for different SNP types

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

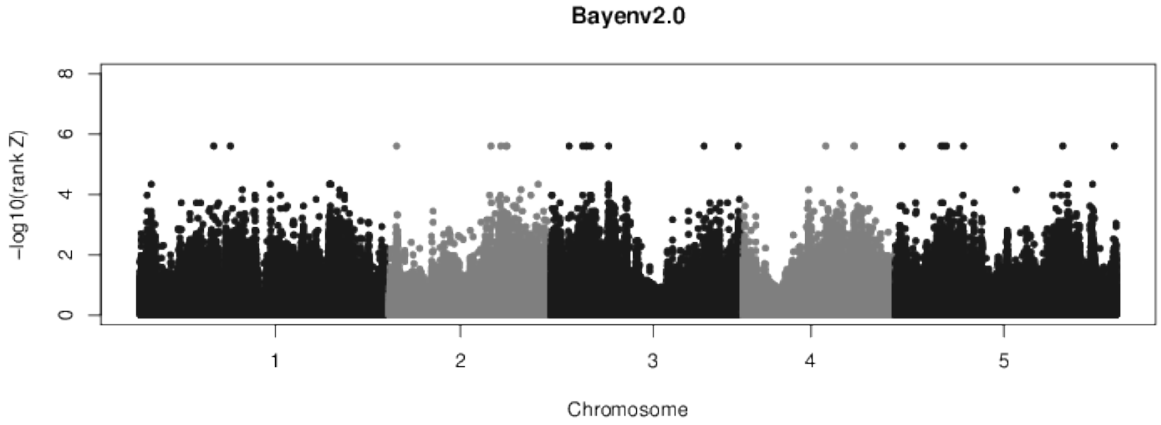

Figure S5: Manhattan plot of the Bayenv results.

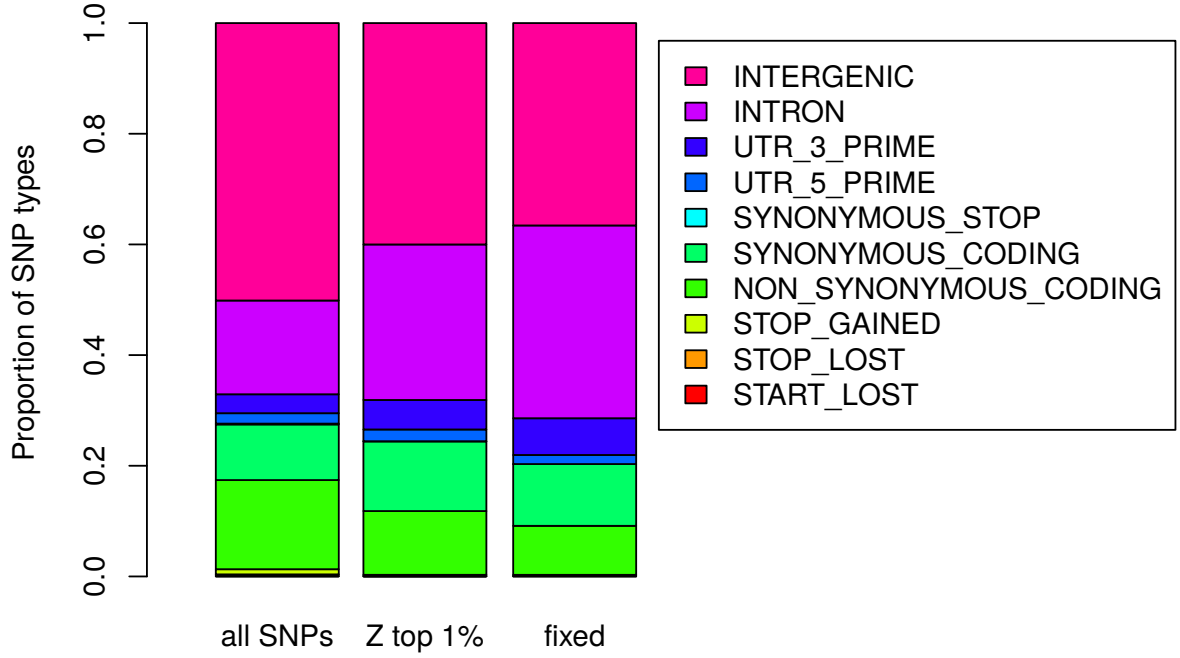

Figure S6: Relative proportion of the different SNP types for all SNPs and three different candidate SNP sets. (Bayenv2.0 top 1% and completely fixed between high and low pops)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

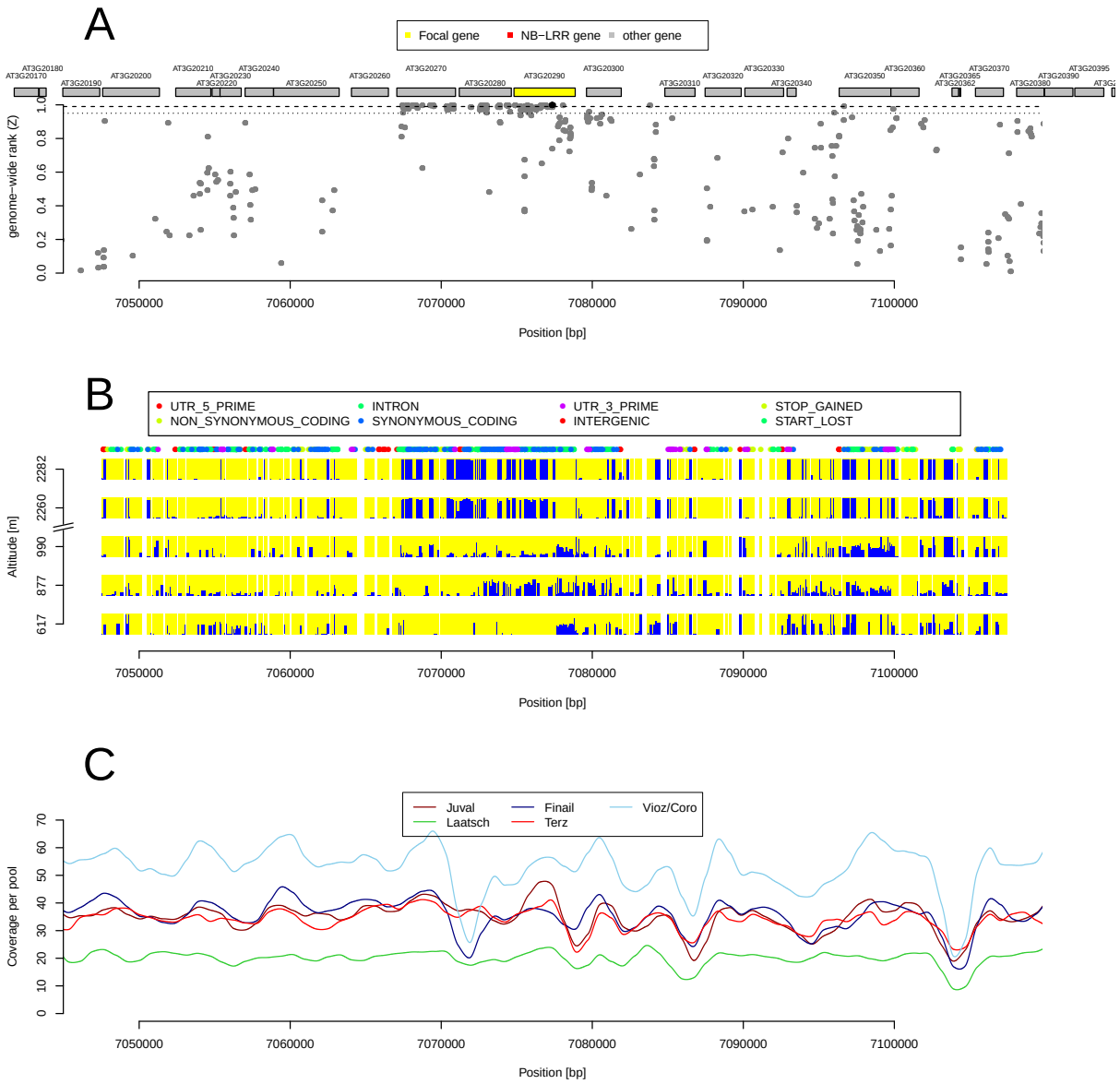

Figure S7: Example of the candidate gene AtEHD1 (AT3G20290) in which allele frequencies were highly differentiated between high and low altitude populations. (A) Plot of the genome-wide relative rank of the Z statistics. High values indicate a strong differentiation. (B) Plot of the relative allele frequencies of the two alleles segregating at a polymorphic site. The height of the bar is proportional to the allele frequency in the pooled sequence data. (C) Sequence coverage of the genomic region in the five different populations.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

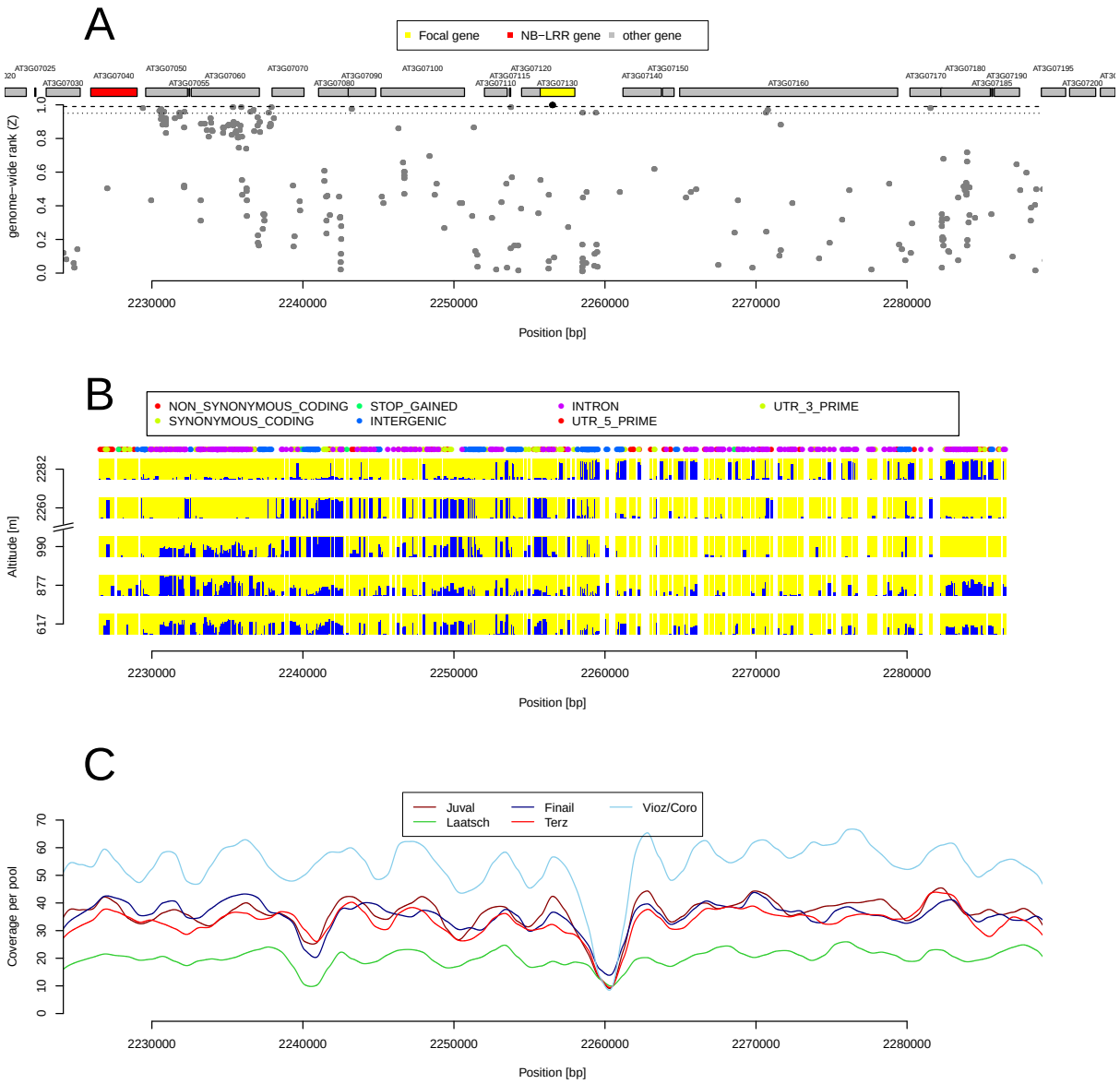

Figure S8: Example of a candidate gene (AT3G07130, encodes the purple acid phosphatase 15) in which allele frequencies differ strongly between the population at the highest altitude and the four lower altitude populations. (A) Plot of the genome-wide relative rank of the Z statistics. High values indicate a strong differentiation. (B) Plot of the relative allele frequencies of the two alleles segregating at a polymorphic site. The height of the par is proportional to the allele frequency. (C) Sequence coverage of the genomic region in the five different populations.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

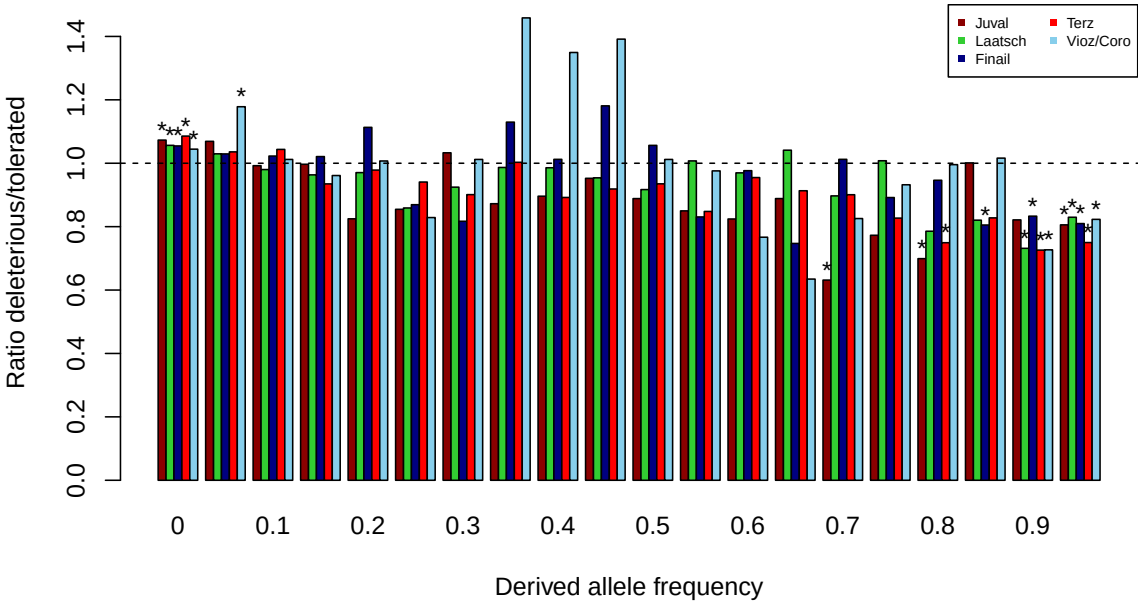

Figure S9: Ratio of relative frequencies of deleterious and tolerated amino acid polymorphisms per frequency bin. If both polymorphism types were neutral, a ratio of 1 (dashed line) would be expected for all frequency classes. Asterisk above bars represent ratios significantly different from 1 (nominal p-value threshold 0.01, tested with a χ^2-test).

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

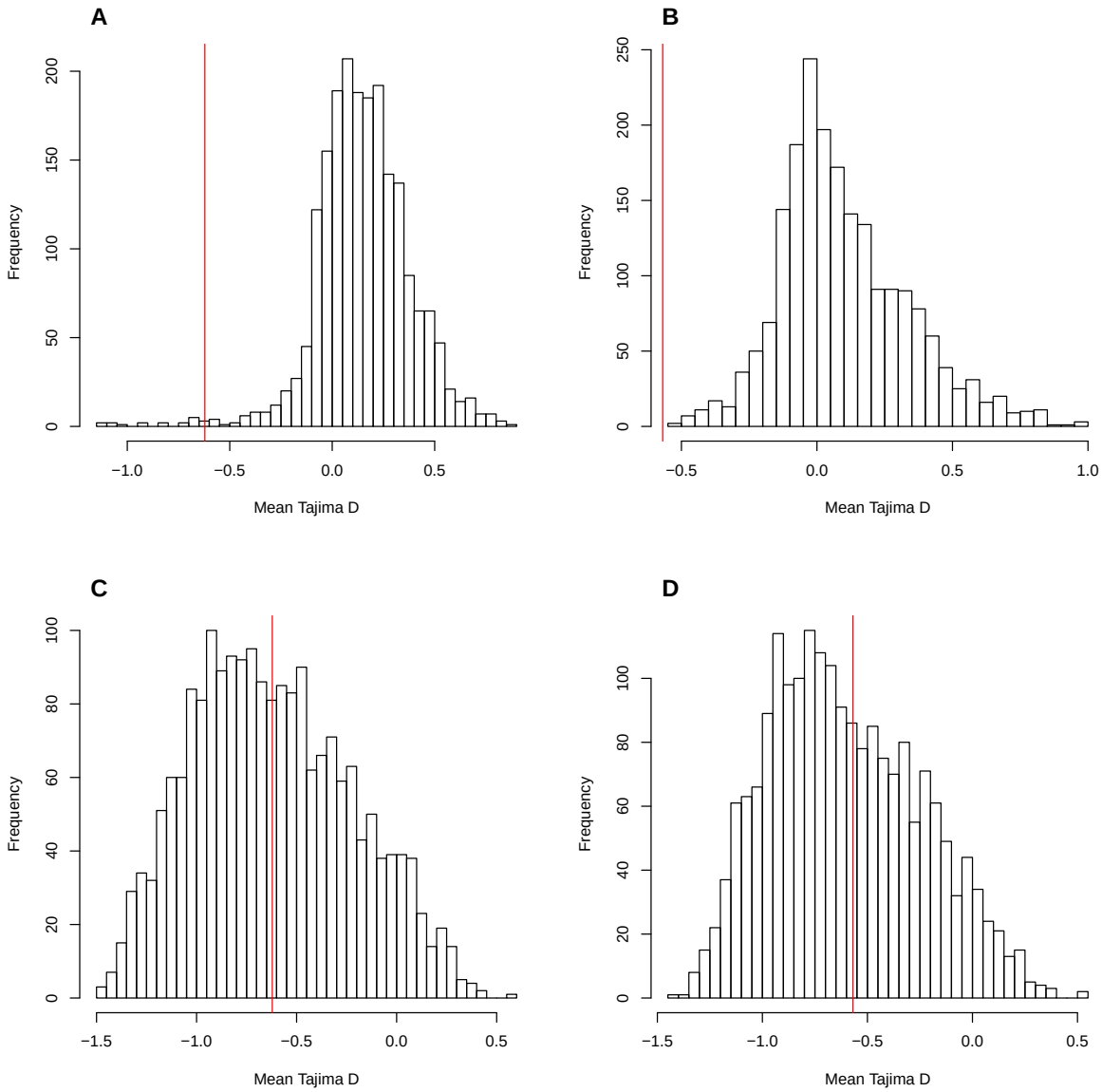

Figure S10: Posterior distribution of Tajima's D values in the high altitude population for the best four ABC models. (A) Juval-Finail stepgrowth model. (B) Terz-Vioz/Coro exponential growth model. (C) Juval-Finail exponential growth with migration model. (D) Terz-Vioz/Coro exponential growth with migration model. Observed values are indicated by the red vertical lines.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Tables

Table S1: Analysis of deviance table of a binomial GLM for dead biomass in response to different dosages of UV-B.

Variable	F-Value$_{df/df}$	P-Value
population	$75_{4/191}$	<0.001
UV-B dosage	$874_{1/190}$	<0.001
population × UV-B dosage	$22_{4/186}$	<0.001

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S2: Average pairwise F_{ST} (lower triangle) and pairwise correlations of Bayenv2.0 correlation matrix among allele frequencies (upper triangle).

	Juval	Laatsch	Finail	Terz	Vioz/Coro
Juval	-	0.79	0.81	0.75	0.67
Laatsch	0.12	-	0.77	0.99	0.71
Finail	0.12	0.12	-	0.71	0.81
Terz	0.12	0.12	0.14	-	0.62
Vioz/Coro	0.14	0.12	0.12	0.16	-

Table S3: Prior distributions of the Juval-Finail step-growth and Terz-Vioz/Coro exponential growth models.

	J-F Stepgrowth		J-F Exponential mig.		T-V/C Exponential		T-V/C Exponential mig.	
Parameter	Min	Max	Min	Max	Min	Max	Min	Max
θ^a	2.8	11.2	2.8	11.2	2.8	11.2	2.8	11.2
ρ^b	0.25	1.02	0.25	1.02	0.25	1.02	0.25	1.02
N_{high}	0.001	1.0	0.001	1.0	0.001	1.0	0.001	1.0
$N^c_{high,growth}$	-	-	0.0	159.7575	0.0	159.7575	0.0	159.7575
$N^d_{high,growth,time}$	-	-	0.011875	0.4975	0.011875	0.4975	0.011875	0.4975
T_{split}	0.0125	0.5	0.0125	0.5	0.0125	0.5	0.0125	0.5
M^e_{12}	-	-	0.0	10.0	-	-	0.0	10.0
M_{21}	-	-	0.0	10.0	-	-	0.0	10.0

[a]Theta is calculated as $4 * N_0 * \mu * L$; N_0=[50,000:200,000]; μ=$7 * 10^{-9}$; L=2000 (length of simulated segments).
[b]Rho is calculated as $4 * N_0 * r * L$; N_0=[50,000:200,000]; r=$0.625 * 10^{-9}$; L=2000 (length of simulated segments).
[c]Growth rate of the high population.
[d]Time of the high population growth.
[e]Migration rate is calculated as $4 * N_0 * m$.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S4: Posterior model probabilities of each demographic scenario calculated by the ABC *postpr* function using the *mnlogistic* method with a 0.005 tolerance to select the best growth model for the Juval-Finail and Terz-Vioz/Coro population pairs.

	Without migration				With migration			
Population pair	No growth	Step	Linear	Exponential	No growth	Step	Linear	Exponential
Juval-Finail	0.000	0.00	0.000	0.000	0.000	0.001	0.000	0.999
Terz-Vioz/Coro	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S5: Soil parameters of surface soil (5-10 cm) in the five population sites.

Population	pH	Organic matter %	Phosphate mg/100g	Potassium mg/100g	Magnesium mg/100g
Juval	6.1	6.17	4.2	21	21
Terz	6.2	4.9	5.5	29	26
Laatsch	6.3	3.47	<1.0	8	8.3
Finail	4.6	17.2	16	58	36
Vioz/Coro	5.2	15.8	22	24	23

Table S6: All 25 SNPs with $Z = 0.5$ in the Bayenv2 analysis.

Chromosome	bp	AGI ID	SNP type	Description
1	9,021,128	AT1G26090	Intron	P-loop containing nucleoside triphosphate hydrolases superfamily protein
1	11,056,544	AT1G31010	Synonymous coding	Organellar single-stranded DNA binding protein 4 (OSB4)
2	902,758	AT2G03060	Intron	*AGAMOUS-LIKE 30*; Pollen development, pollen exine formation
2	12,412,973		Intergenic	
2	13,606,764	AT2G31970	Non-synonymous coding	Encodes the Arabidopsis RAD50 homologue; DNA repair, RNA processing
2	14,276,102	AT2G33760	Synonymous coding	Pentatricopeptide repeat (PPR) superfamily protein
2	14,376,598	AT2G34030	Intron	Calcium-binding EF-hand family protein
3	2,256,537	AT3G07130	Non-synonymous coding	*PURPLE ACID PHOSPHATASE 15*; Pollen germination, seed germination
3	3,897,644	AT3G12220	Intron	*SERINE CARBOXYPEPTIDASE-LIKE 16*; Proteolysis
3	4,298,285	AT3G13290	Non-synonymous coding	Varicose-related (*VCR*)
3	4,427,242	AT3G13560	Non-synonymous coding	O-Glycosyl hydrolases family 17 protein; Carbohydrate metabolic process
3	4,878,614		Intergenic	
3	7,077,355	AT3G20290	Intron	Arabidopsis Eps15 homology domain proteins involved in endocytosis (*AtEHD1*)
3	18,716,502	AT3G50430	Non-synonymous coding	Unknown protein
3	22,868,343	AT3G61780	Non-synonymous coding	Embryo defective 1703 (emb1703); Embryo development ending in seed dormancy
4	10,106,954		Intergenic	
4	13,574,591	AT4G27040	5' UTR	VPS22; Vesicle-mediated transport
4	13,577,662	AT4G27050	Intron	F-box/RNI-like superfamily protein
5	824,047	AT5G03360	Intron	DC1 domain-containing protein; Oxidation-reduction process
5	5,587,807	AT5G17010	Intron	Major facilitator superfamily protein; Transmembrane transport
5	5,903,750	AT5G17860	Non-synonymous coding	Calcium exchanger 7 (*CAX7*)
5	6,252,514		Intergenic	
5	8,335,327		Intergenic	
5	20,429,419		Intergenic	
5	26,727,336	AT5G66930	Intron	Unknown protein

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S7: Enriched GO Biological processes among the top Bayenv2.0 results with nominal p-values ≤ 0.05.

Significant genes	All genes	Description
3	5	**Shade avoidance**, Goslim: response to abiotic or biotic stimulus
4	7	Secondary shoot formation, Goslim: developmental processes
6	12	Leaf vascular tissue pattern formation, Goslim: developmental processes
19	71	**Response to light stimulus**, Goslim: response to abiotic or biotic stimulus
4	10	Negative regulation of transcription, Goslim: transcription
4	12	Cuticle development, Goslim: developmental processes
4	9	**Heat acclimation**, Goslim: response to stress
14	57	Unidimensional cell growth, Goslim: other cellular processes
5	12	**Defense response signaling pathway**, resistance gene-dependent, Goslim: response to stress
8	22	Chloroplast organization, Goslim: cell organization and biogenesis
2	5	**Flavonol biosynthetic process**, Goslim: other cellular processes
11	42	Shoot development, Goslim: developmental processes
4	6	**Phenylpropanoid metabolic process**, Goslim: other cellular processes
5	11	Phloem or xylem histogenesis, Goslim: developmental processes
8	36	**Flower development**, Goslim: developmental processes
3	7	Folic acid and derivative biosynthetic process, Goslim: other cellular processes
4	6	Ovule development, Goslim: developmental processes
23	108	Multicellular organismal development, Goslim: developmental processes
6	24	Plant-type cell wall modification during multidimensional cell growth, Goslim: other cellular processes

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S8: Enriched PO, KEGG, AraCyc categories among the top Bayenv2.0 results with nominal p-values ≤ 0.05.

Significant genes	All genes	Description
9	15	PO:0020144 expressed in apical meristem
14	38	PO:0000229 **expressed in flower meristem**
3	4	**Superoxide radicals degradation**
6	12	**Glucosinolate biosynthesis**
8	20	CDP-diacylglycerol biosynthesis I
8	20	CDP-diacylglycerol biosynthesis II
5	10	glycolipid desaturation
5	11	PO:0020149 expressed in quiescent center
4	9	Glucosinolate biosynthesis from phenylalanine
3	6	Glucosinolate biosynthesis from tetrahomomethionine
3	6	Glucosinolate biosynthesis from trihomomethionine
5	11	Tetrahydrofolate biosynthesis II
10	34	Protein export
3	5	Mevalonate pathway I
3	7	Glucosinolate biosynthesis from homomethionine
3	7	Glucosinolate biosynthesis from hexahomomethionine
3	7	Glucosinolate biosynthesis from pentahomomethionine
2	3	Brassinosteroid biosynthesis II
7	22	PO:0006085 expressed in root meristem
2	2	PO:0007130 expressed during B reproductive growth
3	7	Glucosinolate biosynthesis from dihomomethionine
8	23	α-Linolenic acid metabolism
3	4	**Flavone and flavonol biosynthesis**
4	8	Isoleucine degradation I
3	3	**Anthocyanin biosynthesis** (pelargonidin 3-O-glucoside, cyanidin 3-O-glucoside)
3	8	PO:0000372 expressed in metaxylem
7	19	FormylTHF biosynthesis II

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

28 **Bayenv2**

Table S9: GO terms among top Z signals with nominal p-value$\leq$ 0.05 obtained from an analysis with Gowinda.

Cat	FDR	Genes found	All genes	Description
CC	0.2918	29	125	GO:0009579 thylakoid, Goslim:other intracellular components
MF	0.30478875	3	4	GO:0004784 superoxide dismutase activity, GOslim other enzyme activity
MF	0.30478875	7	14	GO:0008794 arsenate reductase (glutaredoxin) activity, GOslim other enzyme activity
BP	0.30478875	2	2	GO:0016579 protein deubiquitination, Goslim:protein metabolism
BP	0.30478875	3	5	GO:0009641 shade avoidance, Goslim:response to abiotic or biotic stimulus
CC	0.30478875	9	22	GO:0009295 nucleoid, Goslim:other intracellular components
CC	0.30478875	5	10	GO:0005819 spindle, Goslim:other intracellular components
MF	0.30478875	208	1185	GO:0003700 transcription factor activity, GOslim transcription factor activity
BP	0.3569411111	4	7	GO:0010223 secondary shoot formation, Goslim:developmental processes
CC	0.3764346154	6	15	GO:0019005 SCF ubiquitin ligase complex, Goslim:other intracellular components
CC	0.3764346154	50	221	GO:0005783 endoplasmic reticulum, Goslim:ER
BP	0.3764346154	6	12	GO:0010305 leaf vascular tissue pattern formation, Goslim:developmental processes
MF	0.3764346154	17	58	GO:0046872 metal ion binding, GOslim other binding
CC	0.3783357143	8	21	GO:0009524 phragmoplast, Goslim:other cytoplasmic components
CC	0.44608	3	5	GO:0009574 preprophase band, Goslim:other cytoplasmic components
BP	0.493291875	19	71	GO:0009416 response to light stimulus, Goslim:response to abiotic or biotic stimulus
MF	0.5177496	13	56	GO:0016563 transcription activator activity, GOslim other molecular functions
BP	0.5177496	4	10	GO:0016481 negative regulation of transcription, Goslim:transcription

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

BP	0.5177496	4	12	GO:0042335 cuticle development, Goslim:developmental processes
BP	0.5177496	4	9	GO:0010286 heat acclimation, Goslim:response to stress
BP	0.5177496	14	57	GO:0009826 unidimensional cell growth, Goslim:other cellular processes
BP	0.5177496	5	12	GO:0009870 defense response signaling pathway, resistance gene-dependent, Goslim:response to stress
BP	0.5177496	8	22	GO:0009658 chloroplast organization, Goslim:cell organization and biogenesis
MF	0.5177496	6	15	GO:0008017 microtubule binding, GOslim protein binding
BP	0.5177496	2	5	GO:0051555 flavonol biosynthetic process, Goslim:other cellular processes
MF	0.5183137037	13	48	GO:0008415 acyltransferase activity, GOslim transferase activity
BP	0.5183137037	11	42	GO:0048367 shoot development, Goslim:developmental processes
BP	0.5257717857	4	6	GO:0009698 phenylpropanoid metabolic process, Goslim:other cellular processes
BP	0.5337251724	5	11	GO:0010087 phloem or xylem histogenesis, Goslim:developmental processes
BP	0.5614648649	8	36	GO:0009908 flower development, Goslim:developmental processes
MF	0.5614648649	5	10	GO:0016207 4-coumarate-CoA ligase activity, GOslim other enzyme activity
BP	0.5614648649	3	7	GO:0009396 folic acid and derivative biosynthetic process, Goslim:other cellular processes
MF	0.5614648649	5	12	GO:0016291 acyl-CoA thioesterase activity, GOslim hydrolase activity
BP	0.5614648649	4	6	GO:0048481 ovule development, Goslim:developmental processes
BP	0.5614648649	23	108	GO:0007275 multicellular organismal development, Goslim:developmental processes
CC	0.5614648649	148	835	GO:0005634 nucleus, Goslim:nucleus
BP	0.5614648649	6	24	GO:0009831 plant-type cell wall modification during multidimensional cell growth, Goslim:other cellular processes
CC	0.5720152632	2	4	GO:0016272 prefoldin complex, Goslim:cytosol

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

MF	0.5730625641	46	205	GO:0005215 transporter activity, GOslim transporter activity

Table S10: KEGG pathways, PO terms and Aracyc networks among top Z signals with nominal p-value$\leq$ 0.05

Cat	FDR	Genes found	All genes	Description
PO	0.0746	9	15	PO:0020144 expressed in apical meristem
PO	0.133415	14	38	PO:0000229 expressed in flower meristem
ARACYC	0.1537485714	3	4	superoxide radicals degradation
KEGG	0.1537485714	6	12	Glucosinolate biosynthesis
ARACYC	0.1537485714	8	20	CDP-diacylglycerol biosynthesis I
ARACYC	0.1537485714	8	20	CDP-diacylglycerol biosynthesis II
ARACYC	0.1537485714	5	10	glycolipid desaturation
PO	0.1630625	5	11	PO:0020149 expressed in quiescent center
ARACYC	0.235128	4	9	glucosinolate biosynthesis from phenylalanine
ARACYC	0.2970092857	3	6	glucosinolate biosynthesis from tetrahomomethionine
ARACYC	0.2970092857	3	6	glucosinolate biosynthesis from trihomomethionine
ARACYC	0.2970092857	5	11	tetrahydrofolate biosynthesis II
KEGG	0.2970092857	10	34	Protein export
ARACYC	0.324188	3	5	mevalonate pathway I
ARACYC	0.3547026316	3	7	glucosinolate biosynthesis from homomethionine
ARACYC	0.3547026316	3	7	glucosinolate biosynthesis from hexahomomethionine
ARACYC	0.3547026316	3	7	glucosinolate biosynthesis from pentahomomethionine
ARACYC	0.3547026316	2	3	brassinosteroid biosynthesis II
PO	0.3764761905	7	22	PO:0006085 expressed in root meristem
PO	0.3764761905	2	2	PO:0007130 expressed during B reproductive growth
ARACYC	0.4155586957	3	7	glucosinolate biosynthesis from dihomomethionine
KEGG	0.4155586957	8	23	alpha-Linolenic acid metabolism
KEGG	0.42463625	3	4	Flavone and flavonol biosynthesis
ARACYC	0.4529372	4	8	isoleucine degradation I
ARACYC	0.4808680769	3	3	anthocyanin biosynthesis (pelargonidin 3-O-glucoside, cyanidin 3-O-glucoside)
PO	0.4842432609	3	8	PO:0000372 expressed in metaxylem
ARACYC	0.4842432609	7	19	formylTHF biosynthesis II

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S11: Literature collections among top Z signals with nominal p-value≤ 0.05

Cat	FDR	Genes found	All genes	Description
LIT	0.002275	47	159	SOM3-down (TableS4 PubmedID:18545680)
LIT	0.002275	10	14	Changes in the expression levels of genes targeted by microRNA and trans-acting short interfering RNA assayed on ATH1 microarrays
LIT	0.002275	359	1945	genes with lower expression levels in 3d old RBRcs seedlings compared to 3d old wild type seedlings, grown in the presence of 1% sucrose
LIT	0.002275	98	456	Down-regulated expressed in the obe1-1 obe2-2 when compared to wild type (Table S1 PubmedID:19392692)
LIT	0.003648	24	86	up-regulated genes upon infection of Arabidopsis with R. fascians D188
LIT	0.0079742857	268	1459	Up-regulated wt vs ag mutant Stage12 (Table S2 PubmedID:19385720)
LIT	0.0079742857	173	908	Down-regulated 2-fold in WT-Rc50/WT-D (Table S1 PubmedID:19529817)
LIT	0.008375	113	561	ABA Up-regulated Genes (692)
LIT	0.008375	140	707	cluster4 Expression clusters of H3K27me3-enriched genes by K-means clustering.
LIT	0.008375	122	626	Overview of the 810 highly-responsive genes. Additional file 6 lists genes whose log2(ratio) was greater than 1.585 or lesser than -1.585 (corresponding to 3-fold change) in at least one of the MA/M, SA/M or S/M comparisons in the CATMA array experiment.
LIT	0.0084681818	125	606	Induced RNAs to 718 LEC2 (Table S2 PubmedID:16492731)
LIT	0.0094441667	96	474	Preferentially expressed in seed (Table S 7 PubmedID:18539592)
LIT	0.0138346154	234	1272	Up-regulated 2-fold in YHB-D /WT-D (Table S1 PubmedID:19529817)
LIT	0.0139393333	323	1766	genes with higher expression levels in 3d old RBRcs seedlings compared to 3d old wild type seedlings, grown in the presence of 1% sucrose
LIT	0.0139393333	290	1588	genes with lower expression levels in 3d old RBRcs seedlings compared to 3d old wild type seedlings, grown in the absence of sucrose

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.016344375	230	1275	Genes Deregulated in det3 under Restrictive Conditions (Nitrate)
LIT	0.0182861111	270	1430	genes with higher expression levels in 3d old RBRcs seedlings compared to 3d old wild type seedlings, grown in the absence of sucrose
LIT	0.0182861111	133	679	Up-regulated genes Inferred from Ratios r4 (Table S2 PubmedID:16789830)
LIT	0.0196794737	123	673	down-differentially regulated by Pieris brassicae oviposition on wild-type, coi1-1, and sid2-1 plants
LIT	0.0229915	150	806	Up-regulated genes Inferred from Ratios r3 (Table S2 PubmedID:16789830)
LIT	0.0262841667	18	61	Up-regulated (1.5-fold change, p = 0.01) by silicon supply in Arabidopsis plants infected with Erysiphe cichoracearum (Table S1 PubmedID:17082308)
LIT	0.0262841667	417	2474	¿2x, vs. Sav+resc., root+seedl.
LIT	0.0262841667	73	330	Down-regulated carpel-expressed genes that were used to enhance the prediction of carpel-specific transcripts by ag in the predictions of array elements of the oligonucleotide array representing floral organ-expressed transcripts are summarized. (Table S12 PubmedID:15100403)
LIT	0.0262841667	199	1097	Down-regulated 2-fold in YHB-Rc50/WT-D (Table S1 PubmedID:19529817)
LIT	0.0268456	99	489	Up-regulated in the stn7-1 psad1-1 mutant compared to Col-0. (Table S5 PubmedID:19706797)
LIT	0.0269688462	30	119	Decreased genes in TOC1-ox compared with WT (Table S1 PubmedID:19816401)
LIT	0.0269855172	23	86	Genes significantly changing in at least two of three conditions: downregulated in jaw-D leaves or apices, upregulated in rTCP4:GFP plants (logit-T p¡0.05 and common variance ¿ 2 fold). Promoters were arbitrarily defined as 800 to 10 bp from ATG.
LIT	0.0269855172	364	2116	Differentially expressed genes using a p-value ¡ 0.05 which are early seed specific (ESS) in the GHL Data. (Table S2 PubmedID:18923020)
LIT	0.0269855172	14	41	genes repressed in stop1-mutant, and down-regulated or stable in WT with Al treatment

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.0274623333	53	239	Up-regulated carpel-expressed genes that were used to enhance the prediction of carpel-specific transcripts by ap2 in the predictions of array elements of the oligonucleotide array representing floral organ-expressed transcripts are summarized. (Table S12 PubmedID:15100403)
LIT	0.0279632258	63	300	pklpkr2 UP List of genes deregulated in pkl and pkl pkr2 roots
LIT	0.0296978125	46	197	Up-regulated carpel-expressed genes that were used to enhance the prediction of carpel-specific transcripts by ap1 in the predictions of array elements of the oligonucleotide array representing floral organ-expressed transcripts are summarized. (Table S12 PubmedID:15100403)
LIT	0.0319874286	190	1017	Up-regulated 2-fold in YHB-Rc50/WT-D (Table S1 PubmedID:19529817)
LIT	0.0319874286	142	819	Fold induction relative to T0 of genes significantly differentially expressed upon auxin treatment
LIT	0.0319874286	277	1559	Genes differentially regulated in the SCR microarray time course (Table S2 PubmedID:20596025)
LIT	0.0382271795	22	84	100 galls distinctive DEG with the lowest Fc List of the co-expresed DEG 3 days after infection of Meloidogyne javanica(MJ) between GCs and galls and the galls distinctive genes
LIT	0.0382271795	108	575	Down-regulated genes Inferred from Ratios r5 (Table S2 PubmedID:16789830)
LIT	0.0382271795	7	16	dn,Expression of antioxidant network genes following the transition to darkness
LIT	0.0382271795	8	17	dn,aox1a treated vs aox1a normal,Relative transcript abundance for genes encoding antioxidant defence components located in Arabidopsis
LIT	0.03849925	91	439	dn,Microarray analysis of Wild type vs 35S:ZAT10 line 14 (OE)
LIT	0.0430307317	226	1270	Up-regulated wt vs ag mutant after bolting (Table S2 PubmedID:19385720)
LIT	0.0448921429	29	123	Increase in both 2HS IP and 9HS IP (Table S4 PubmedID:18665916)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.0493845455	199	1106	Down-regulated 2-fold in YHB-D /WT-D (Table S1 PubmedID:19529817)
LIT	0.0493845455	57	267	DELLA-repressed (DELLA-down) genes in the imbibed ga1-3 seeds
LIT	0.0511073333	27	101	UP enhanced regulation in control Genes differentially expressed in response to drought in 35S:ABF3 and control plants
LIT	0.0539280435	61	298	DN differentially expressed genes in AtNudt7-1 versus Wild-type plants growing in 12:3:1 potting mix
LIT	0.0565732653	42	167	Dwon-regulated carpel-expressed genes that were used to enhance the prediction of carpel-specific transcripts by ap3 in the predictions of array elements of the oligonucleotide array representing floral organ-expressed transcripts are summarized. (Table S12 PubmedID:15100403)
LIT	0.0565732653	10	28	Down-regulated in sdg4 mutant flowers (Table 1 PubmedID:18252252)
LIT	0.0565732653	13	39	Up regulated (¿ 2 fold) with drought in WT and C2 (Table S3 PubmedID:18849493)
LIT	0.0623896552	122	671	Up-regulated genes Inferred from Ratios r2 (Table S2 PubmedID:16789830)
LIT	0.0623896552	362	2064	Upregulated genes in hyl1-2. (Table S3 PubmedID:19633021)
LIT	0.0623896552	35	155	Down-regulated in the stn7-1 mutant compared to Col-0. (Table S3 PubmedID:19706797)
LIT	0.0623896552	117	592	Nuclear RP Up Unambig (Table S2 PubmedID:20675573)
LIT	0.0623896552	185	1031	Downregulated genes between TEV and TEV-at infected plants (TableS3 PubmedID:18545680)
LIT	0.0623896552	94	517	Up-regulated genes Inferred from Ratios r1 (Table S2 PubmedID:16789830)
LIT	0.0623896552	125	648	EP-ESS generated from the list of differentiall expressed genes using p -value ¡ 0.05. (Table S5 PubmedID:18923020)
LIT	0.0623896552	80	406	downregulated in the imbibed ga1-3 seeds (GA-upregulated or GA-up)
LIT	0.0623896552	169	952	Down-regulated wt vs ag mutant after bolting (Table S2 PubmedID:19385720)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.0632477966	49	220	Down-regulated genes from meta-analysis (q¡0.05) by GL3 DM 24hrs (Table S 2 PubmedID:19247443)
LIT	0.0637395	28	118	Decrease of effect of pkl and of uniconazole on transcript level of gene. (Table S 5 PubmedID:18539592)
LIT	0.063917541	117	611	significantly differential expression patterns between wild-type, mpk4 and 35S-MKS1 plants
LIT	0.0705620968	6	12	dn,Expression of antioxidant network genes that do not have a clear cellular localization following transition to darkness
LIT	0.0713777273	38	175	Up-regulated carpel-expressed genes that were used to enhance the prediction of carpel-specific transcripts by pi in the predictions of array elements of the oligonucleotide array representing floral organ-expressed transcripts are summarized. (Table S12 PubmedID:15100403)
LIT	0.0713777273	88	452	Up-regulated in brm-5/essp3 Leaves (Table S1 PubmedID:18508955)
LIT	0.0713777273	12	40	Up-regulated genes (top 50)in the albino mutant (Table S2 PubmedID:17468106)
LIT	0.0713777273	352	2063	Venn analysis of differentially expressed genes; down-regulated genes (TableS4b PubmedID:18400103)
LIT	0.0718035211	211	1184	Down-Genes differentially regulated in dms4 compared to drd1. (Table S3 PubmedID:20010803)
LIT	0.0718035211	300	1745	dn,Significant genes changing during CaLCuV infection
LIT	0.0718035211	10	30	Repressed SSTF Class 3 genes that are Repressed in WT-R1, WT-Rc and/or pifq-D relative to WT-D. (Dataset4 PubmedID:19920208)
LIT	0.0718035211	16	67	Upregulated -AGL24 -genes by microarray analysis (Table S1 PubmedID:18339670)
LIT	0.0718035211	77	382	UP Differentially regulated genes in hsp70-15 knockout plants.
LIT	0.0718127778	9	26	Downregulated in shoot between [NH4NO3 +CO(NH2)2]- and [NH4NO3]-Arabidopsis grown plants (reference). (TableS5 PubmedID:18508958)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.0720809091	115	619	Up-regulated 2-fold in WT-Rc15/WT-D (Table S1 PubmedID:19529817)
LIT	0.0720809091	146	851	Down-regulated by WT -N vs. WT FN (Dataset 2 PubmedID:19933203)
LIT	0.0720809091	122	618	Up-6x2-List of genes that are upregulated or downregulated in both the platforms. (Table S3 PubmedID:20406451)
LIT	0.0720809091	335	2016	Down-regulated expressed genes of Agrobacterium tumefaciens(P¡0.01 all)-induced Arabidopsis tumors versus tumor-free inflorescence stalk tissue. (Table S3 PubmedID:17172353)
LIT	0.0720809091	146	852	Up-regulated by WT FN vs. WT -N (Dataset 1 PubmedID:19933203)
LIT	0.0830333333	124	674	Up-regulated at least 2-fold after a shift from growth light (100 mol photons m-2 s-1) to 3 h high fluence blue light (200 mol photons m-2 s-1) (Table S2 PubmedID:17478635)
LIT	0.0853624051	10	31	LED 7 (up; all tissues) in root tissue cluster (Arabidopsis) (TableS4 PubmedID:15918878)
LIT	0.086574875	21	98	up-regulated by the indole-3-acetic acid treatment.
LIT	0.0872672289	8	24	up,aox1a normal vs Col-0 normal,Relative transcript abundance for genes encoding antioxidant defence components located in Arabidopsis
LIT	0.0872672289	11	33	down-Genes responsive to WUS induction
LIT	0.0872672289	30	130	Down-regulated genes in uninfected gh3.5-1D versus wild-type plants (Table S2 PubmedID:17704230)
LIT	0.0912563529	23	97	genes that are induced in Ler but not hy5-1:
LIT	0.0912563529	351	2112	Down-regulated in Agrobacterium tumefaciens-induced tumors of Arabidopsis thaliana. (Table S2 PubmedID:17172353)
LIT	0.0928162791	56	280	Up-regulated at least 2-fold in flu versus wild type, ex1/flu versus wild type, ex2/flu versus wild type, and ex1/ex2/flu versus wild type (Data Set 1 PubmedID:17540731)
LIT	0.0930332184	16	64	dn,Genes differentially expressed in P35S:PAR1-GG seedlings

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.09601875	145	781	Up-regulated 2-fold in WT-Rc50/WT-D (Table S1 PubmedID:19529817)
LIT	0.0972077528	89	461	Induced by auxin during lateral root initiation in tissues surrounding the initiation site (Table S1 PubmedID:18622388)
LIT	0.1028495604	124	662	Down-differentially regulated in WS-2 compared to Col-O. (Table S9 PubmedID:19706797)
LIT	0.1028495604	8	20	Upregulated commonly in the gra-D mutants and located outside of the duplications. (Table S2 PubmedID:19508432)
LIT	0.1089384783	40	170	Up-genes were selected if they were detected as differentially expressed upon both the Dex and Dex+Chx treatments but not upon the Chx treatment (Table S5 PubmedID:20360106)
LIT	0.1124013978	44	208	genes that are induced in Col but not cop1-4:
LIT	0.1124185106	20	75	Up-regulated-WT-temperature(Sample 29C / Sample 20C) (Table S1 PubmedID:19686536)
LIT	0.1176825263	262	1557	Genes Deregulated in det3 under Restrictive Conditions (Temperature)
LIT	0.11889	5	12	TF more induced-Differential gene expression analysis of hypoxic stress AND cell types (DatasetsS7 PubmedID:19843695)
LIT	0.1219590722	9	27	Down- regulated Secretory Pathway Genes (SPGs) with modulated levels of transcripts in 5- and 11-day-old hypocotyls (File 4 PubmedID:19878582)
LIT	0.1265104902	37	178	Genes Induced Maximally by Glucose at 2 Hours (218)
LIT	0.1265104902	95	527	Genes Deregulated in det3 under Permissive Conditions
LIT	0.1265104902	355	2113	OST generated from the list of differentiall expressed genes using p -value ¡ 0.05. (Table S5 PubmedID:18923020)
LIT	0.1265104902	7	20	Down-expression levels of glucosinolate-metabolism and primary sulfur-metabolism genes in response to MP (Table 3 PubmedID:17189325)
LIT	0.1265104902	81	413	Up-regulated by pnp1-1 +P vs WT +P in three hours (Table S3 PubmedID:19710229)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.1278679612	5	9	Significantly co-up-regulated in dcl1-7, dcl4-2, and rdr6-15 (Table S 1 PubmedID:16129836)
LIT	0.1289255238	51	246	Repressed expressed genes upon SA treatment. (TableS 1 PubmedID:17496105)
LIT	0.1289255238	64	326	Up-fis1x2-List of genes that are upregulated or downregulated in both the platforms. (Table S3 PubmedID:20406451)
LIT	0.1291354717	120	681	Down-regulated 2-fold in WT-Rc15/WT-D (Table S1 PubmedID:19529817)
LIT	0.1294504673	31	141	Induced-3h-NO3 (Table S 6 PubmedID:16299223)
LIT	0.1295172222	31	160	Affymetrix expression data validated by qRT-PCR
LIT	0.1296294737	16	59	Diff-The number of unique hits per million are shown for Arabidopsis genes greater than 1000 bases long in wild type shoots. A value of 0 indicates that, at the level of sequencing depth used in this work, no reads aligning uniquely to that particular transcript were found. (Table S2 PubmedID:20512117)
LIT	0.1296294737	4	10	Induced-group in Slightly later in the diurnal cycle ¿2-fold changes within 30 min after adding 15 mM Suc to C-starved seedlings was extracted from Osuna et al. (2007). (Table II PubmedID:18305208)
LIT	0.1296294737	200	1174	The organ-expressed genes identified with the oligonucleotide array are listed (Table S7 PubmedID:15100403)
LIT	0.1296294737	71	374	Microarray analysis of 35S:DREB2A CA plants. Genes with fold change of ¿2 are listed.–fluorescence intensity of each cDNA of 35S:DREB2A CA/fluorescence intensity of each cDNA of vector control line.
LIT	0.1296294737	389	2267	Proteins identified by MS/MS in isolated wt and ppi2 plastids
LIT	0.1296294737	16	59	Up-regulated 24 hours after infiltration of M. persicae saliva and their expression frofile after M. persicae feeding. (Table S1 PubmedID:19558622)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.1300405085	5	14	Down-regulated genes of the GO category response-to-salt-stress in hmgb1 (Table S1 PubmedID:18822296)
LIT	0.1300405085	22	103	Upregulated genes-Overlapping between the present study and that of Himanen et (2004) with the cluster combination and the cluster number and profile, respectively (Table S 4 PubmedID:16243906)
LIT	0.1300405085	70	372	Upregulated in BRZ Col (Table S4 PubmedID:18599455)
LIT	0.1300405085	23	102	IAA up-regulated genes induction not affected 2 fold or more up or down by by simultaneous glucose treatment
LIT	0.13463675	19	86	ARR1SR BA15 vs ARR1SR BA0 down,Lists of all significantly regulated genes in the wild type and line ARR1-S-8
LIT	0.13463675	53	263	Down-regulated genes -spl/nzz (TableS1 PubmedID:17905860)
LIT	0.1411063636	19	77	pklpkr2 UP H3K27me3 List of genes deregulated in pkl and pkl pkr2 roots
LIT	0.1429948361	245	1424	differently regulated –cell cycle regulation during the growth process of leaves 1 and 2 of Arabidopsis
LIT	0.1433647967	20	78	Genes suggested to be involved in the transition from a vegetative state to flowering and their predicted regulators
LIT	0.1447189516	55	260	Depressed by KIN10 markedly overlaps with that induced by starvation conditions and is antagonized by increased sugar availability (600 genes). (Table S4 PubmedID:17671505)
LIT	0.1495619841	40	186	Increased genes in TOC1-ox compared with WT (Table S1 PubmedID:19816401)
LIT	0.1495619841	42	202	DELLA-upregulated (DELLA-up) genes in the ga1-3 young flower buds
LIT	0.14981	12	46	Down-regulated genotype (elf3) x temperature of 23 interaction (p¡0.01, uncorrected) (Table S4 PubmedID:19187043)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.1501121875	202	1194	Down-regulated expressed genes of Agrobacterium tumefaciens (P¡0.01 + functional categories)-induced Arabidopsis tumors versus tumor-free inflorescence stalk tissue. (Table S3 PubmedID:17172353)
LIT	0.1509633333	7	22	dn,aox1a treated vs Col-0 treated,Relative transcript abundance for genes encoding antioxidant defence components located in Arabidopsis
LIT	0.1509633333	158	950	Downregulated genes in hyl1-2. (Table S5 PubmedID:19633021)
LIT	0.1509633333	45	221	Genes Repressed maximally by Glucose at 6Hours (275)
LIT	0.1509633333	90	480	ARR1SR vs Col-0 down,Lists of all significantly regulated genes in the wild type and line ARR1-S-8
LIT	0.1512332331	127	732	Downregulated Se-reponsive genes in root tissue (Table 1A PubmedID:18251864)
LIT	0.1527363433	36	161	Repressed-4h-carbon fixcation (Table S 6 PubmedID:16299223)
LIT	0.1554525899	10	30	Down-Coregulated Genes in stn7-1, psad1-1, and psae1-3 Leaves (Table 3 PubmedID:19706797)
LIT	0.1554525899	6	20	common ABA-responsive genes that reduced in both srk2d/e/i and areb1/areb2/abf3 triple mutants in comparison to WT after 6 h of treatment with 50 ?M ABA
LIT	0.1554525899	11	43	dn,Diurnal gating of cold-responsive TFs.the difference between the expression attained after morning cold treatment at ZT2 (Cm) versus evening cold treatment at ZT14 (Ce)
LIT	0.1554525899	50	260	Genes Induced Maximally by Glucose at 6 Hours (343)
LIT	0.1554525899	8	23	Down regulated at least 4 times in jaw-1D plants (Table 2 PubmedID:12931144)
LIT	0.1564065	331	2005	Up-Genes differentially regulated in dms4 compared to drd1. (Table S3 PubmedID:20010803)
LIT	0.1564184507	76	398	Genes showing more than twofold reduced transcript level in 35S:AtMYB44 Arabidopsis treated with 250 mM NaCl

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.1564184507	57	304	down-regulated in 14 day-old leaves of Arabidopsis haf2-1 mutant as assayed by CATMA microarray
LIT	0.1584254545	69	374	Light up-regulated in Root (TableS6 PubmedID:15888681)
LIT	0.167433125	13	55	Singificance Up¿4.0FC (S 02 PubmedID:15668208)
LIT	0.1676214857	40	219	Up-regulated by AZC (Dataset1 PubmedID:19244141)
LIT	0.1676214857	27	131	Genesrepressed (DOWN) in leaves of long-term experiment : raw and normalized values and flags are reported. Results for the three replicates are reported.
LIT	0.1676214857	103	588	up,genes that are consistently affected in triple mutant pollen
LIT	0.1676214857	23	107	Suppressed by ABA in rop10-1 (Table S 4 PubmedID:16258012)
LIT	0.1676214857	85	465	Down-expressed after SLWF nymph feeding (Table S2 PubmedID:17189325)
LIT	0.1676214857	8	30	Genes with specific expression in light-treated seedlings.
LIT	0.1676214857	26	124	Genes Positively Regulated by HY5 (Table 1 PubmedID:17337630)
LIT	0.1676214857	33	161	Repressed-3h-mannitol (Table S 6 PubmedID:16299223)
LIT	0.1676214857	24	126	List of genes up-regulated 2 h after treatment with IAA
LIT	0.1676214857	17	72	up,Microarray identification of genes differentially expressed in wrky33-1 mutants as compared to wild-type
LIT	0.1676214857	102	552	Down-regulated genes -ap3es (TableS1 PubmedID:17905860)
LIT	0.1676214857	69	368	OST-ESS generated from the list of differentiall expressed genes using p -value ¡ 0.05. (Table S5 PubmedID:18923020)
LIT	0.1676214857	9	33	Down regulated (¿ 2 fold) with drought in WT and C2 (Table S5 PubmedID:18849493)
LIT	0.1676214857	34	163	dn,Overview of MeJA-responsive genes(methyl JA)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.1676214857	3	6	Repressed by infections of PPV and other positive sense RNA viruses (Table S8 PubmedID:18613973)
LIT	0.1676214857	47	252	Down-regulated expressed in the obe1i obe2i mutant seedlings when compared to wild type (Table S2 PubmedID:19392692)
LIT	0.1676214857	26	121	Up-regulated in 4hPT v 0.5hPT (Table S11 PubmedID:19714218)
LIT	0.1676214857	120	707	Genes differentially expressed in response to Sclerotinia sclerotiorum(S-sclerotiorum) infection in leaves of in wild-type and coi1 mutant
LIT	0.1676214857	74	398	double/wt triple/wt DN genes significantly affected in either the double(agl66 agl104-2), triple(agl65 agl66 agl104-1), or quadruple(agl65 agl66 agl94 agl104-1) mutant pollen.
LIT	0.1676214857	29	134	UPR up-regulated genes dependent on bZIP28 function or expression
LIT	0.1676214857	104	595	Up-regulated expression (oler rapa FDR average CommVarsu) between the diploid progenitors B. rapa and B. oleracea (Dataset S1 PubmedID:19274085)
LIT	0.1676214857	28	142	Up-regulated genes -35S::atNF-YB1 (TableS3 PubmedID:17923671)
LIT	0.1676214857	88	506	Down-regulated by COL vs atAF bgh (Table 2 PubmedID:18694460)
LIT	0.1676214857	4	9	Up-regulated specifically in flu after paraquat treatmenta (Table S3 PubmedID:14508004)
LIT	0.1676214857	14	57	Co-expressed with higher Fc values in GCs than in galls List of the co-expresed DEG 3 days after infection of Meloidogyne javanica(MJ) between GCs and galls and the galls distinctive genes

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.1676214857	119	678	Up- regulated expressed Arabidopsis thaliana (Ws-2) genes Suppl-Figure 2A: Table lists all genes shown in Figure 2A differentially transcribed in the following experiments: (i) Three hours post inoculation of wounded inflorescence stalks with A. tumefaciens strain C58 (C58 3 hpi) and (ii) with strain GV3101 (GV3101 3 hpi); (iii) six days post inoculation with strain C58 (C58 6 dpi) and (iv) with GV3101 (GV3101 6 dpi), as well as (v) 35 day-old tumors induced by strain C58 (tumor 35 dpi). (DatasetsS 1 PubmedID:19794116)
LIT	0.1676214857	2	3	Decreased in vte2 relative to wild-type at 3 days. (Table S2 PubmedID:17194769)
LIT	0.1676214857	64	382	down-Genes differentially regulated by Pieris brassicae oviposition
LIT	0.1676214857	13	45	Induced transcripts(top 50) by ethylene (Table S3 PubmedID:16920797)
LIT	0.1676214857	19	95	OtsB DN Genes more than 2-fold affected compared with the wild type
LIT	0.1676214857	6	18	Up-expression levels of glucosinolate-metabolism and primary sulfur-metabolism genes in response to EC (Table 3 PubmedID:17189325)
LIT	0.1685826136	143	841	UP WUSCHEL(WUS) REGULATION SCORES for WUS response genes.
LIT	0.1732157303	56	277	Significantly regulated -uncapped predicted miRNA (termed miSVM, Lindow et al., 2007) targets. Only genes with uncapped/total mRNA ratio ¿ 2 and a P value ¡ 0.01 in at least one time point during early flower development are shown. (Tables4 PubmedID:18952771)
LIT	0.1732157303	285	1674	Down-[ga1] vs [Col-0II] (Table S2 PubmedID:20844019)
LIT	0.1759167222	92	514	Up-Further analysis indicated 640 genes increased, their transcript levels by 4-fold or more (p ? 0.001) as result of the dehydration stress. (Table S3 PubmedID:21050490)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.1759167222	44	232	Genes induced (UP) in the medium-term experiment : raw and normalized values and flags are reported. Results for the three replicates are reported.
LIT	0.1779594737	45	216	Down-regulated in the stn7-1 psae1-3 mutant compared to WS-2. (Table S7 PubmedID:19706797)
LIT	0.1779594737	23	114	IAA up-regulated genes up-regulated by glucose treatment alone.
LIT	0.1779594737	70	371	Opposite regulation (induced in BL/H3BO3, repressed in GCs) co-expresed and distinctive DEG between differentiating vascular cell suspensions (BL/H3BO3) and GCs.
LIT	0.1779594737	37	187	Down-regulated in ahg2-1 (log2 ¡ -1 or log2 ¿1) (Table S1 PubmedID:19892832)
LIT	0.1779594737	28	133	pkl UP List of genes deregulated in pkl and pkl pkr2 roots
LIT	0.1779594737	6	16	Differentially regulated -Targeted by NA small RNAs (Table S1 PubmedID:17400893)
LIT	0.1779594737	7	20	Down-regulated and auxin-related genes in tengu-transgenic plants as identified by microarray analysis (Table 1 PubmedID:19329488)
LIT	0.1779594737	10	39	Reduced RNA level genes with at least a 5-fold in tdt-1 mutants (Table S1 PubmedID:17513503)
LIT	0.1779594737	22	92	Down-regulated in the psae1-3 mutant compared to WS-2 (Table S6 PubmedID:19706797)
LIT	0.1779594737	346	2108	EP generated from the list of differentiall expressed genes using p -value ¡ 0.05. (Table S5 PubmedID:18923020)
LIT	0.1779876963	36	167	ZAT12 dn,Overrepresentation of the CBF2 and ZAT12 regulons among selected gene lists
LIT	0.1801860104	7	21	Up-miRNA targets significantly up-regulated at least two-fold in dcl1 early globular embryos relative to wild-type (Table S1 PubmedID:21123653)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.1801860104	35	155	Down-regulated carpel-expressed genes that were used to enhance the prediction of carpel-specific transcripts by pi in the predictions of array elements of the oligonucleotide array representing floral organ-expressed transcripts are summarized. (Table S12 PubmedID:15100403)
LIT	0.1814145641	4	10	up,Complete list of genes diferentially expressed in leaves of tpc1-2 versus WT
LIT	0.1814145641	23	116	down-regulated by the indole-3-acetic acid treatment.
LIT	0.1827628141	102	584	Up-regulated in the psad1-1 mutant compared to Col-0. (Table S4 PubmedID:19706797)
LIT	0.1827628141	2	3	Up regulated with drought (¿2 fold) in InsP 5-ptase plants (Table S4 PubmedID:18849493)
LIT	0.1827628141	54	307	down-regulated drl1-2/Ler
LIT	0.1827628141	26	125	Down -regulated genes displaying differential expression between mature and immature lateral nectaries - MLN/ ILN . (Table S9 PubmedID:19604393)
LIT	0.1838005366	15	60	Up- regulated expressed *Arabidopsis thaliana* (Ws-2) genes Supplemental Figure 4 that are affected either by C58 or GV3101 at 6hpi (DatasetsS 1 PubmedID:19794116)
LIT	0.1838005366	34	165	N-terminal acetylated peptides(N-ace-peptid) identified by MS/MS in Toc159cs leaves
LIT	0.1838005366	24	110	Aza-dC+TSA downregulated (TableS1 PubmedID:15516340)
LIT	0.1838005366	7	29	Prominent upregulated genes in steroid-inducible MKK3DD plants
LIT	0.1838005366	28	129	The 5% most highly expressed genes in mature Arabidopsis trichomes
LIT	0.1838005366	7	22	Greatest up-regulated genes by 1.0 ?m IAA at 3h (Table 5a PubmedID:15086809)
LIT	0.184001699	9	35	Down Regulation of Cold/Light-Responsive Genes at Least Two-Fold (P-value less than 0.05 at least in one condition)
LIT	0.184435782	5	15	Induced during PPV infection but either induced or suppressed by infections of other positive sense RNA viruses (Table S8 PubmedID:18613973)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.184435782	117	684	downregulated genes in the ga1-3 young flower buds (GA-upregualted or up)
LIT	0.184435782	97	533	Down-regulated genes -ms1 (whole inflorescence) (TableS1 PubmedID:17905860)
LIT	0.184435782	57	312	Down-2x6-List of genes that are upregulated or downregulated in both the platforms. (Table S3 PubmedID:20406451)
LIT	0.184435782	54	289	DELLA-repressed (DELLA-down) geens in the ga1-3 young flower buds
LIT	0.1864415566	55	311	down-regulated elo2/Ler
LIT	0.1883802347	60	313	upregulated genes in the ga1-3 young flower buds (GA-downregulated or GA-down)
LIT	0.1911436449	39	221	up,Differentially expressed genes (548) that responded to 35 ?M NAE12:0 treatment in 4-day old Arabidopsis seedlings
LIT	0.1916938889	23	120	down-regulated genes upon infection of Arabidopsis with R. fascians D188
LIT	0.1916938889	10	38	Induced by both M. persicae feeding (De Vos et al. 2005) and saliva infiltration (Table 1 PubmedID:19558622)
LIT	0.1919147465	6	16	Down-Genes affected in pye-1 mutants by Fe.–Microarry analysis of wild-type and pye-1 mutants after 24 hours of exposure to +Fe or Fe media. (Table S3 PubmedID:20675571)
LIT	0.1941900889	24	113	Upregulated BBM target genes identified as significantly in all three statistical analyses (ANOVA, SAM, t-test) (TableS2 PubmedID:18663586)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.1941900889	127	742	Differentially regulated by 1.5 fold (P ¡0.05) in the wild-type plants by chitooctaose 30 minutes after treatment as revealed by the comparison between the chitooctaose-treated and water treated samples. The expression pattern of these genes in the similarly treated mutant plants was also included after the wild-type Data. WT: wild-type Col-0; Mu: the atLysM RLK1 mutant; 8mer: treatment with chitooctaose; water: treatment with distilled water; I, II, and III: biological replicates. False Discovery Rate (Number) of 200 permutations: Median: 0.0% (0), and 90th percentile: 99.9% (889). (Table 1 PubmedID:18263776)
LIT	0.1941900889	3	8	Up-regulated on exhibit uniconazole-P-dependent based on the criteria (Pkl vs. Wt p ¡ 0.05 ?? Upkl vs. Uwt p ¡ 0.05) and for which the corresponding transcript level is elevated two-fold or more (Table S2 PubmedID:12834400)
LIT	0.1941900889	12	43	Up-regulated genes in MYB34 (Table S4 PubmedID:18829985)
LIT	0.1941900889	78	403	Decreased-regulation by KIN10 (1021 genes). (Table S3 PubmedID:17671505)
LIT	0.1941900889	78	403	Down-regulated-KIN10 (Table S6 PubmedID:18305208)
LIT	0.1941900889	111	647	Jasmonate-esponsive genes (22hrs)in stamens of Arabaidopsis thaliana (Table S1 PubmedID:16805732)
LIT	0.1941900889	8	27	Repressed-group in Start of the extended night ¿2-fold changes within 30 min after adding 15 mM Suc to C-starved seedlings was extracted from Osuna et al. (2007). (Table II PubmedID:18305208)
LIT	0.1963785903	55	293	D - Genes repressed by Cd 15 and 30 M (447),Genes significantly induced or repressed exclusively by Cd2+ 30 M or by both Cd2+ 15 M and 30 M

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.1963785903	119	661	Down-regulated in the stn7-1 psad1-1 mutant compared to Col-0. (Table S5 PubmedID:19706797)
LIT	0.2000454852	297	1832	Up-regulated expressed genes of Agrobacterium tumefaciens(P¡0.01 all)-induced Arabidopsis tumors versus tumor-free inflorescence stalk tissue. (Table S3 PubmedID:17172353)
LIT	0.2000454852	250	1511	Up-significantly altered expression (Dark Fold Change) in 35S:MIF1 seedlings (false discovery rate ¡0.1; MIF1, MINI ZINC FINGER 1) (Table S1 PubmedID:16412086)
LIT	0.2000454852	68	372	DN Differentially regulated genes in hsp70-15 knockout plants.
LIT	0.2000454852	45	233	Down-regulated Gene List-Cold (Table S 6 PubmedID:16214899)
LIT	0.2000454852	45	235	Up-2x6-List of genes that are upregulated or downregulated in both the platforms. (Table S3 PubmedID:20406451)
LIT	0.2000454852	121	681	UP msh1-recA3 transcripts specifically regulated by a certain mitochondrial impairment.
LIT	0.2000454852	133	778	Up-regulated at least 2-fold after a shift from growth light (100 mol photons m-2 s-1) to 3 h 1000 mol photons m-2 s-1 in wild type (Table S1 PubmedID:17478635)
LIT	0.2000454852	78	440	DN msh1-recA3 transcripts specifically regulated by a certain mitochondrial impairment.
LIT	0.2000454852	11	45	Negative expression of list of non-redundant genes that were differentially expressed comparing flowering mutants to their near isogenic controls (TableS2 PubmedID:15908439)
LIT	0.2000454852	31	157	Significant down-regulation in 72hrs in response to iron-deprivation (Table S2 PubmedID:18436742)
LIT	0.2009856574	23	101	up,Genes showing significant change in expression level between wild type and transgenic lines
LIT	0.2009856574	24	108	Cold Down-regulated Genes Whose Expression is Affected by ice1 (Table S 13 PubmedID:16214899)
LIT	0.2009856574	81	459	upregulated in the imbibed ga1-3 seeds (GA-downregulated or GA-down)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.2009856574	9	37	Commonly down regulated (¿ 2 fold) with drought (161 transcripts) (Table S8 PubmedID:18849493)
LIT	0.2009856574	6	20	Induced columella markers at regeneration 22 hrs compared to regeneration 0 hrs (Table S2 PubmedID:19182776)
LIT	0.2009856574	11	39	Up-regulated in CandiDate BABY BOOM target genes in DEX + CHX-treated 35S::BBM:GR seedlings as compared to DEX + CHX-treated wild-type seedlings (Table2 PubmedID:18663586)
LIT	0.2009856574	19	88	All Genes That Pass Significance Criteria following Induction of GLK1.
LIT	0.2009856574	61	325	Up-regulated genes that were nonadditively expressed in allopolyploid line 1200 relative to the 1:1 parent mix–parentmix 5250 FDR CommonVar (Dataset S5 PubmedID:19274085)
LIT	0.2009856574	7	22	Down-regulated by Sulfur and SLIM1 (Table S1 PubmedID:17114350)
LIT	0.2009856574	3	8	up,Genes up-regulated and down-regulated in 9-h dehydrated ahk1 mutant
LIT	0.2009856574	9	32	Late repressed TFs in Jasmonate responsive transcription factors (Table S3 PubmedID:16805732)
LIT	0.2009856574	136	783	Down-regulated genes Inferred from Ratios r1 (Table S2 PubmedID:16789830)
LIT	0.2009856574	31	163	Up-regulated carpel-expressed genes that were used to enhance the prediction of carpel-specific transcripts by ap3 in the predictions of array elements of the oligonucleotide array representing floral organ-expressed transcripts are summarized. (Table S12 PubmedID:15100403)
LIT	0.2009856574	27	124	Up regulated by BR late (between 3 and 24 hours) (Table S 1 PubmedID:15681342)
LIT	0.2033985714	14	63	Arabidopsis genomic regions down-regulated expression upon rrp4est mutation (TableS2 PubmedID:18160042)
LIT	0.2037863636	2	3	dn,Differentially expressed genes in jai3-1 vs wild-type plants , after JA treatment

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.2044906693	33	176	Repressed in the vfb mutants compared to the wild type (Table S2 PubmedID:17435085)
LIT	0.2057408627	246	1466	Induced in PPV-infected Arabidopsis leaf tissues 17 days post inoculation (Table S1 PubmedID:18613973)
LIT	0.2134856202	126	730	Down-6x2-List of genes that are upregulated or downregulated in both the platforms. (Table S3 PubmedID:20406451)
LIT	0.2134856202	70	415	Significant up-regulation in 72hrs in response to iron-deprivation (Table S2 PubmedID:18436742)
LIT	0.2134856202	112	631	down-Genes with altered expression above 1.3 lower bound fold change in bol-D young leaves
LIT	0.2138952672	5	15	dn,aox1a normal vs Col-0 normal,Relative transcript abundance for genes encoding antioxidant defence components located in Arabidopsis
LIT	0.2138952672	41	223	Repressed at Least 1.5-Fold in the Crab Shell Chitin Mixture-Treated Seedlings (Table S 5 PubmedID:15923325)
LIT	0.2138952672	59	310	Down-regulated genes that were nonadditively expressed in allopolyploid line 1200 relative to the 1:1 parent mix–parentmix 5250 FDR CommonVar (Dataset S5 PubmedID:19274085)
LIT	0.2138952672	41	210	Down-regulated (LU) Gene List-Cold Late (Table S 9 PubmedID:16214899)
LIT	0.2139394361	16	72	Up in 2-6, bound (by fdr) (Table S3 PubmedID:20675573)
LIT	0.2139394361	7	35	Aza-dC downregulated (TableS1 PubmedID:15516340)
LIT	0.2139394361	9	32	Strong upregulation of transcripts from infected roots of plants –clubroot disease
LIT	0.2139394361	7	21	up,gl3-3-trichome,The 25 most up- and down-regulated genes in gl3-3 and try-JC trichomes
LIT	0.2145473606	42	221	genes that are induced in Ler and hy5-1:
LIT	0.2145473606	90	481	Up-regulated genes that were nonadditively expressed in allopolyploid line 1200 relative to the 1:1 parent mix–parentmix 1250 FDR PerGeneVar (Dataset S3 PubmedID:19274085)

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIT	0.2145473606	8	35	All Genes That Pass Significance Criteria following Induction of GLK2.

bioRxiv preprint doi: https://doi.org/10.1101/023051; this version posted July 22, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

DevelopmentCoreTeam (2014) *R: A Language and Environment for Statistical Computing*. Vienna.