ETH Library

The genomic basis of adaptation to calcareous and siliceous soils in Arabidopsis lyrata

Journal Article

Author(s):
Guggisberg, Alessia; Liu, Xuanyu; Suter, Léonie; Mansion, Guilhem; Fischer, Martin C.; Fior, Simone; Roumet, Marie; Kretzschmar, Ruben; Koch, Marcus A.; Widmer, Alex

Publication date:
2018-12

Permanent link:
https://doi.org/10.3929/ethz-b-000312391

Rights / license:
In Copyright - Non-Commercial Use Permitted

Originally published in:
Molecular Ecology 27(24), https://doi.org/10.1111/mec.14930

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

The genomic basis of adaptation to calcareous and siliceous soils in *Arabidopsis lyrata*

Alessia Guggisberg[1*], Xuanyu Liu[1], Léonie Suter[1], Guilhem Mansion[1], Martin C. Fischer[1], Simone Fior[1], Marie Roumet[1], Ruben Kretzschmar[2], Marcus A. Koch[3], Alex Widmer[1]

[1]ETH Zurich, Institute of Integrative Biology, Universitätstrasse 16, CH-8092 Zürich. [2]ETH Zurich, Institute of Biogeochemistry and Pollutant Dynamics, Universitätstrasse 16, CH-8092 Zurich. [3] Heidelberg University, Centre for Organismal Studies Heidelberg, Im Neuenheimer Feld 345, D-69120 Heidelberg. [*]Author for correspondence: Alessia Guggisberg, E-mail: alessia.guggisberg@env.ethz.ch

Abstract

Edaphic conditions are important determinants of plant fitness. While much has been learnt in recent years about plant adaptation to heavy-metal contaminated soils, the genomic basis underlying adaptation to calcareous and siliceous substrates remains largely unknown. We performed a reciprocal germination experiment and whole-genome re-sequencing in natural calcareous and siliceous populations of diploid *Arabidopsis lyrata* to test for edaphic adaptation and detect signatures of selection at loci associated with soil-mediated divergence. In parallel, genome scans on respective diploid ecotypes from the *A. arenosa* species complex were undertaken, to search for shared patterns of adaptive genetic divergence. Soil ecotypes of *A. lyrata* display significant genotype-by-treatment responses for seed germination. Sequence (SNPs) and copy-number variants (CNVs) point towards loci involved in ion transport as the main targets of adaptive genetic divergence. Two genes exhibiting high differentiation among soil types in *A. lyrata* further share trans-specific single-nucleotide polymorphisms with *A. arenosa*. This work applies experimental and genomic approaches to study edaphic adaptation in *A. lyrata* and suggests that physiological response to elemental toxicity and deficiency underlies the evolution of calcareous and siliceous ecotypes. The discovery of shared adaptive variation between sister species indicates that ancient polymorphisms contribute to adaptive evolution.

Keywords: copy-number variants, divergent selection, edaphic adaptation, reciprocal common garden transplants, trans-specific polymorphism, whole-genome re-sequencing

Introduction

During their evolutionary history, plants have recurrently diversified in response to heterogeneous ecological conditions (Givnish, 2010; Waser & Campbell, 2004) by developing adaptations to climatic, topographic and edaphic discontinuities (Kruckeberg, 1986; Mason, 1946; Rajakaruna, 2004). Chemical factors such as elemental composition are known to constitute important drivers for the evolution of soil ecotypes (Beadle, 1953; Rajakaruna & Boyd, 2008), as exemplified by a wealth of studies on heavy metal tolerance (Krämer, 2010), resistance to aluminium toxicity (Ryan et al., 2011) and adaptation to serpentine soils (Turner, Bourne, Von Wettberg, Hu, & Nuzhdin, 2010). Edaphic adaptation may also originate following differential responses to physical (*e.g.* water holding capacity)

and biotic (*e.g.* microbiome) properties of different soil types (Rajakaruna, 2017). However, the genetic basis underlying the divergence between edaphic ecotypes adapted to a large spectrum of soil acidity remains largely unknown (*e.g.* Gould, McCouch, & Geber, 2014, 2015).

Since the beginning of the -omics era, most integrative efforts have been deployed towards a better understanding of plant adaptation to heavy-metal rich soils (Arnold et al., 2016; Hanikenne et al., 2013; Hanikenne et al., 2008; Turner et al., 2010; Turner, Von Wettberg, & Nuzhdin, 2008; Wright et al., 2015) or to extreme salt levels (Dassanayake et al., 2011; Eshel et al., 2017; Wu et al., 2012). Yet, these soils cover a limited range of existing edaphic conditions and impose steep selective clines at study sites (*e.g.* at the edge of mine tailings). It currently remains open whether genetic changes underlying edaphic adaptation in these systems are representative for adaptation to more gradual differences. From an analytical point-of-view, most studies have focused on single-nucleotide polymorphisms (SNPs) and ignored copy-number variants (CNVs), despite their known importance in organismic evolution (Kondrashov, 2012; Żmieńko, Samelak, Kozłowski, & Figlerowicz, 2014) and edaphic adaptation (Hanikenne et al., 2013; Hanikenne et al., 2008; Maron et al., 2013; Suryawanshi et al., 2016; Wright et al., 2015).

Zonal soils are determined by the local climate (*e.g.* Podzols or Chernozems) and generally form well-defined (mature) horizons. By contrast, azonal soils exhibit simple (immature) profiles, which are mainly determined by non-climatic factors such as bedrock type or relief. For example, Leptosols developing on calcareous (calcite/dolomite-rich) sedimentary rocks (hereafter calcareous soils) differ from those derived from siliceous (quartz/feldspar-rich) igneous or metamorphic rocks (hereafter siliceous soils) with respect to chemical composition, physical properties, and pH, which in turn largely determines nutrient availability and elemental toxicity in these soils (Blume et al., 2010). Calcareous soils are rather porous and neutral-to-alkaline (pH 7-8), and exhibit high concentrations of plant-available calcium, but deficiencies in potassium, sulphate, phosphate and iron. By contrast, siliceous soils are rather impermeable, acidic (pH 4-5) and rich in exchangeable aluminium, but depleted in exchangeable base cations such as calcium and magnesium.

Theophrastus in the fourth century BC (Hughes, 1985) noticed the contrasting vegetation found between calcareous and siliceous soils, raising the attention of botanists (Braun-Blanquet, 1964; Ellenberg, 1958; Unger, 1836) and plant physiologists (Bothe, 2015; Kinzel, 1983; Lee, 1999) thereafter to edaphic plant adaptation. The recurrent observation of significant genotype-by-environment (GxE) effects, when reciprocally growing calcicole and calcifuge species (*i.e.*, species that preferentially occur on calcareous or siliceous substrates, respectively) under controlled edaphic conditions, suggests that soil chemistry underlies the evolution of edaphic specialists (Clymo, 1962; Davies & Snaydon, 1973a, 1973b). Biotic interactions, including competition among species, may further promote adaptive divergence, because ruderal and stress-tolerant calcicole species are typically displaced by more competitive calcifuge species in densely vegetated siliceous environments (Gigon, 1971, 1987; Tansley, 1917). In turn, adverse physical conditions such as drought may exacerbate the selective pressures in calcareous habitats, thereby favouring calcicole over calcifuge ecotypes (Bamberg & Major, 1968; Grime & Curtis, 1976; Michalet, Gandoy, Joud, & Pagès, 2002).

The present work investigates the genomic basis of adaptation to calcareous and siliceous substrates in *Arabidopsis lyrata* subsp. *petraea* (L.) O'Kane & Al-Shehbaz (hereafter *A. lyrata*). This species comes with a fully sequenced and annotated genome (Hu et al., 2011; Rawat et al., 2015), grows on diverse soil types, including several different kinds of azonal soils (Clauss & Koch, 2006), is an obligate outcrosser with high standing genetic variation

(Clauss & Mitchell-Olds, 2006), and has been used as a model for the evolution of edaphic adaptation to serpentine soils (Arnold et al., 2016; Turner et al., 2010; Turner et al., 2008). To identify shared polymorphisms among sister lineages, we also study the outcrossing *A. arenosa* species complex (hereafter *A. arenosa*; Hohmann et al., 2014; Koch, Wernisch, & Schmickl, 2008; Kolář et al., 2016; Novikova et al., 2016; Schmickl, Paule, Klein, Marhold, & Koch, 2012). Both species have evolved diploid edaphic ecotypes growing on calcareous and siliceous soils.

A germination experiment performed to assess local adaptation in *A. lyrata* reveals significant genotype-by-treatment responses among soil ecotypes, indicating edaphic adaptation at early life stages. The genomic signatures of divergent selection point towards soil chemistry as the main driver of adaptive genetic differentiation between calcareous and siliceous populations and highlight the potential importance of CNVs in edaphic adaptation. The high convergence in functional categories and partial overlap between candidate gene sets reveal that *A. lyrata* and *A. arenosa* have evolved similar strategies to cope with contrasting edaphic conditions and have in part recruited the same trans-specific polymorphisms for edaphic adaptation. Owing to the combination of multiple analytical approaches, this study presents an integrative picture of edaphic adaptation in *A. lyrata*.

Materials and Methods

Plant material. During spring 2012, we collected both silicagel-dried leaf tissues and living plants of *A. lyrata* in seven natural populations from calcareous (n = 4) and siliceous (n = 3) sites spanning the Bohemian Massif, the Czech Republic and the East Austrian Forealps (Fig. 1a, Supporting Information Table S1). During the following summer, we further sampled silicagel-dried leaf tissues of *A. arenosa* in seven natural populations from calcareous (n =3) and siliceous (n = 4) sites (Supporting Information Fig. S1, Table S1). Permission for sampling *A. arenosa* in the Tatra National Park was granted by the Regional Environmental Office in Prešov (permit no. 2012/1155-6031/KM-R).

Soil analyses. Soil pH was measured in the field with the HELLIGE® Soil Reaction pH Tester Kit (Ben Meadows, WI, USA), and the presence of carbonates ($CaCO_3$) was assessed by the degree of effervescence of CO_2 after addition of a few drops of 10% HCl to both rock and soil samples from the A horizon.

Three topsoil samples from each *A. lyrata* locality were further investigated in the lab. Soil pH was measured after equilibrating 10 g of soil with 25 mL 0.01 M $CaCl_2$ using a Mettler Toledo MP220 pH Meter (Mettler-Toledo GmbH, Greifensee, Switzerland). Total elements (except carbon and nitrogen) were quantified by X-ray fluorescence (XRF) spectrometry (Spectro-X-Lab 2000, Spectro, Kleve, Germany) of pressed pellets made of 4 g powdered, air-dried soil mixed with 0.9 g of wax (C-Wachs, Hoechst, Frankfurt, Germany). Carbon and nitrogen contents were determined by combustion of 1-2 mg powdered, air-dried soil on a LECO 932-CHNS elemental analyser (Leco, Krefeld, Germany).

Germination experiment. Between October 2012 and September 2013, controlled within-population crosses were performed in the greenhouse, to obtain seeds for germination in Petri dishes containing either siliceous or calcareous soil during summer 2014. Each type of soil was collected at the original population sites, air-dried and sieved through 2 mm grids, before being mixed by soil type in equal volume parts. Petri dishes were then filled up with 20 ml

dried soil mixed with either 14 ml (siliceous substrate) or 12 ml distilled water (calcareous substrate), to ensure comparable water level across treatments.

Eight seeds from four independent crosses were sown per population on the respective soil type and placed in a dark climate chamber set to 4° C for 16 days. The same approach was repeated with reciprocal crosses of the same parents. Subsequently, Petri dishes were transferred to a climate chamber set for 16 h daylight, 20° C and 70% humidity. The germination date (defined as the appearance of both cotyledons) was recorded on a daily basis for a total of 896 seeds: 7 populations × 4 parent pairs × 2 reciprocal cross directions × 8 seeds × 2 treatments (Supporting Information Fig. S2).

The effect of seed origin and soil treatment on days to germination were examined with linear mixed effects models using the function *lmer*, while the germination success of seeds originating from either soil type was assessed with binomial logit models using the function *glmer* with the R package *lme4* (Bates, Maechler, Bolker, & Walker, 2014). Seed origin, treatment and their interactions were considered as fixed effects, and original population, reciprocal cross, maternal parent and petri dish as random effects. To meet model assumptions, days to germination were log-transformed. Significance of fixed effects was assessed by model comparisons using likelihood ratio tests. All graphical outputs were obtained using R 3.1.1 (R Development Core Team, 2018).

Sequencing and mapping of pooled sequence data. For each population of *A. lyrata* and *A. arenosa*, high-quality genomic DNA of 30 individuals was extracted using the DNeasy Plant Mini Kit (Qiagen, Basel, Switzerland) and pooled into equimolar quantities for 100 bp paired-end whole-genome re-sequencing (Futschik & Schlötterer, 2010; Rellstab, Zoller, Tedder, Gugerli, & Fischer, 2013; Schlötterer, Tobler, Kofler, & Nolte, 2014) on an Illumina HiSeq 2000 platform, with an insert size of 500 bp and an average coverage of 100-130× per population. All sequence data were filtered for adapters, low-quality bases (Phred score < 20) and singletons using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). True-Seq library preparation and Illumina sequencing were performed by the Quantitative Genomics Facility (D-BSSE, ETH Zurich, Basel, Switzerland).

Trimmed paired-end reads from each population pool were mapped against the published reference genome (excluding organellar DNA) of the North American *A. lyrata* subsp. *lyrata* strain MN47 (Hu et al., 2011) using BWA-ALN (H. Li & Durbin, 2009), allowing four and seven mismatches for *A. lyrata* and *A. arenosa*, respectively. Using SAMtools (H. Li et al., 2009), putative PCR duplicates were finally removed and only read pairs with a minimum mapping quality of MAPQ 20 were retained.

Adaptive model. Since *Arabidopsis* species diverged very recently (around the onset of the Pleistocene glaciations; Hohmann, Wolf, Lysak, & Koch, 2015), we searched the genome for signs of selective sweeps that are characteristic of rapid adaptive evolution. The genomic analyses combined in this study aimed at detecting consistent signals resembling hard sweeps that result from the rapid fixation of beneficial mutations in response to positive selection in all calcareous-siliceous comparisons (Cutter & Payseur, 2013; Flood & Hancock, 2017; Messer & Petrov, 2013). Accordingly, we integrated results from multiple approaches, ranging from SNP- and window-based analyses to CNV detection methods, to derive a comprehensive picture of edaphic adaptation in *A. lyrata*. The SNP outlier dataset alone may, however, include examples of soft sweeps, because multiple haplotypes carrying the same adaptive SNP may have been retained (Hermisson & Pennings, 2005).

Detection of divergent SNPs in pooled sequence data. For SNP calling, alignments were converted to mpileup files using SAMtools (H. Li et al., 2009) and synchronised using PoPoolation2 (Kofler, Pandey, & Schlötterer, 2011), keeping only positions with a minimum base quality of 20, a minimum coverage of 20 and a maximum coverage of twice the median coverage of each population. To further minimise SNP calling errors, the minimum count of the minor allele was set to eight, repetitive regions were removed, and all indels including two flanking nucleotides were masked.

Highly differentiated outlier SNPs that may result from selection were either identified by contrasting the genetic differentiation (F_{ST}) of individual SNPs to the empirical genome-wide distribution of F_{ST} (Akey, Zhang, Zhang, Jin, & Shriver, 2002; de Mita et al., 2013), or by Cochran-Mantel-Haenzsel (CMH) tests, which consider calcareous-siliceous population pairs as replicates in the statistical analyses, as implemented in PoPoolation2 (Kofler et al., 2011). In the first case, positions that showed F_{ST}-values greater than the upper 95% quantile calculated upon the entire empirical distribution in all (12) possible calcareous-siliceous comparisons were retained as outliers. Because CMH tests more readily identify allele frequency differences in regions with high coverage (Kofler et al., 2011), we reduced the coverage of every position to a minimum of 20 for these analyses, by subsampling the alleles while retaining the exact fraction in PoPoolation2 (Kofler et al., 2011). A SNP was considered differentiated if the estimated p-value after Bonferroni correction was lower than 0.001 across all possible combinations. In contrast to traditional F_{ST} cut-off approaches, CMH tests constitute a complementary method to detect subtle but consistent changes in allele frequencies.

Genes containing outlier SNPs were screened for over-represented gene ontology (GO) terms using the R package SNP2GO (Szkiba, Kapun, von Haeseler, & Gallach, 2014) to infer the adaptive processes at play in *A. lyrata*. These genes were further retained for subsequent comparisons using Venn diagrams. SNP2GO was run with default settings on GO annotations (derived from TAIR homologs) of *A. lyrata* (v384) downloaded from the Phytozome v12.0 database (Rawat et al., 2015). Only GO terms with an FDR-corrected p-value lower than 0.01 were considered. Enriched GO terms were summarised with REVIGO (Supek, Bosňjak, Škunca, & Šmuc, 2011).

Detection of genomic islands of divergence in pooled sequence data. To further characterise sweep signals in *A. lyrata*, we searched the genome for islands of divergence that are characterised by both increased F_{ST} and between-population divergence D_a (also known as D_m or π_D) across non-overlapping 100-SNP stepping windows. The relative measure D_a allows one to detect signatures of positive selection by assessing the local increase of absolute divergence (D_{xy} or π_B) between ecotypes with respect to within-deme nucleotide diversity (Charlesworth, 1998; Cruickshank & Hahn, 2014; Muir, Bergero, Charlesworth, & Filatov, 2011). Window-based analyses correct for large-scale heterogeneity along chromosomes that may indirectly affect the threshold used to detect outlier loci. The need to set a fixed window size however comes with the risk of averaging out small-scale patterns of genomic divergence.

F_{ST} values for each given window were obtained using PoPoolation2 (Kofler et al., 2011), while D_a estimates (Nei & Li, 1979) were calculated between any two populations x and y as: $D_a = D_{xy} - (\pi_x + \pi_y)/2$, by approximating the nucleotide diversity (π; Ferretti, Ramos-Onsins, & Pérez-Enciso, 2013) per population and the average number of nucleotide substitutions (D_{xy}; Nei & Li, 1979) between populations using R (R Development Core Team, 2018). The window size was chosen to reflect the amount of variation present per gene

(median: 122 SNPs), while simultaneously ensuring a high ratio of true to false positives (Beissinger, Rosa, Kaeppler, Gianola, & de Leon, 2015).

A cubic smooth spline was generated for each chromosome (scaffold_1 to scaffold_8) using the highest smoothing parameter (spar = 1) in R (R Development Core Team, 2018) and all window-based estimates at hand. For each regression line, we then computed 95% confidence intervals from the standard error in jackknife residuals. Windows that consistently showed F_{ST} and D_a estimates greater than the upper 95% quantile in at least 11 of the 12 calcareous-siliceous comparisons were extracted as putative windows undergoing divergent selection. All graphs were produced using R 3.3.1 (R Development Core Team, 2018).

We complemented our window-based analyses by analysing 70% consensus sequences from each population using the software Saguaro (Zamani et al., 2013). Saguaro constitutes a model-free approach that combines a hidden Markov model with a self-organising map to reconstruct local topologies (*viz.* cacti) among aligned genomes. Specifically, the algorithm searches genomic regions with similar patterns of differentiation and assigns them to hypothesised phylogenies. Saguaro was run for 20 iterations with a transition penalty of 130 on aligned, IUPAC ambiguity coded consensus sequences of scaffold_1 to scaffold_8.

Detection of copy-number variants in individual sequence data of *A. lyrata*. To assess whether CNVs are associated with nucleotide-based signals of selective sweeps in *A. lyrata*, 21 individuals (three per population) were sequenced (see Sequencing and mapping of pooled sequence data) for an average coverage of 15× per individual. After quality trimming and filtering for PCR duplicates with ConDeTri (Smeds & Künstner, 2011), each sequence library was mapped against the reference genome (Hu et al., 2011) using the default settings in BWA-MEM (H. Li, 2013), retaining only read pairs with a minimum mapping quality of MAPQ 20 in SAMtools (H. Li et al., 2009).

Deletions (DEL) and tandem duplications (TD) were detected using three complementary sequencing-based approaches that differ with regard to size spectrum, power and resolution: (i) the read depth (RD) approach implemented in CNVnator (Abyzov, Urban, Snyder, & Gerstein, 2011) with different bin sizes set for different coverage ranges (100 bp for > 15× coverage, 200 bp for 10-15× coverage, 300 bp for < 10× coverage); (ii) the discordant read pair (RP) approach implemented in DELLY (Rausch et al., 2012) with default settings; and (iii) the split read (SR) approach implemented in Pindel (Ye, Schulz, Long, Apweiler, & Ning, 2009) with the maximum size of structural variations (–x option) set to 129,472 bp.

The basic idea underlying the RD approach is that the density (depth) of mapped reads within a genomic region is roughly proportional to the copy number of that region (Xi, Lee, & Park, 2012). The RD approach is suited for estimating absolute copy number, but suffers from low sensitivity for small CNVs (<1 kb) and low breakpoint resolution (Bellos, Johnson, & Coin, 2012). The RP approach infers CNVs by assessing the discordance of mapped read pairs from the expected insert size and orientation (Alkan, Coe, & Eichler, 2011). Although the RP approach outperforms the RD approach in terms of resolution and size range, it suffers from false positive calls when mapping is unreliable as in repetitive regions (W. Li & Olivier, 2013). The SR approach finally relies on the principle that reads that overlap with CNV breakpoints can be split into two fragments, to infer the position of CNVs (Ye et al., 2009).

To combine the CNV calls from all three approaches across all 21 *A. lyrata* genomes, we performed a so-called 'precision-aware merging' (Mills et al., 2011; Zichner et al., 2013). Confidence intervals around the breakpoints of CNVs were taken from Zichner et al. (2013)

for each approach (RD approach: 1 kb outwards, 400 bp inwards; RP approach: 50 bp outwards, 250 bp inwards; SR approach: 10 bp outwards, 10 bp inwards). CNV calls with overlapping confidence intervals at both start and end coordinates were considered as the same variant and were merged into a final dataset, keeping only CNVs that were detected with at least two approaches.

The genotype for each DEL and TD was inferred with DELLY (Rausch et al., 2012), which counts all normal and discordant pairs at the predicted breakpoints to infer one of three possible genotypes: homozygous reference (RR), heterozygous variant (RA) and homozygous variant (AA). In cases where the software failed to clearly distinguish heterozygous from homozygous variants, the genotype was defined as 'ambiguous'. With the genotype information at hand, we identified calcareous- and siliceous-specific DEL and TD by calculating the allele frequency differences between soil ecotypes, i.e. by assessing the absolute difference in the presence/absence of given CNVs between individuals from calcareous and siliceous populations, respectively. To avoid inaccurate estimations, CNVs presenting more than two ambiguous genotypes (i.e. more than 5% missing data) were excluded. Python was used to assign CNVs to genomic elements they overlapped with.

Assessment of shared alleles with *A. arenosa*. To assess whether putative adaptive alleles found in *A. lyrata* have been recurrently recruited in closely related species, we sequenced and analysed seven diploid populations of the *A. arenosa* species complex (Hohmann et al., 2014; Koch et al., 2008; Kolář et al., 2016; Schmickl et al., 2012), including three populations from calcareous (*A. carpatica* nom. prov.) and four from siliceous (*A. neglecta* (Schultes) O'Kane & Al-Shehbaz) bedrock for F_{ST} and CMH SNP outliers (see Detection of divergent SNPs in pooled sequence data and Supporting Information Table S1). Yet, we refrained from performing window-based or phylogenetic analyses within *A. arenosa* for assessment of edaphic divergence, because calcareous and siliceous populations belong to two distinct taxonomic entities at low and high elevations in the Carpathians, such that putative signals of edaphic divergence are potentially confounded with signals from demographic changes in general and altitudinal divergence in particular (Hohmann et al., 2014; Koch et al., 2008; Kolář et al., 2016; Schmickl et al., 2012). Yet, to identify trans-specific polymorphisms, we searched both *A. lyrata* and *A. arenosa* consensus sequences for shared SNPs among ecotypes using Saguaro (see "Detection of genomic islands of divergence in pooled sequence data").

Gene tree reconstructions. To assess the history of root expressed high-affinity sulphate transporter 1;1 (*SULTR1;1;* homolog AL6G42430 of AT4G08620) at the genus level, 62 published *Arabidopsis* genomes, covering all diploid taxa sampled throughout their range, were investigated phylogenetically for this region (Novikova et al., 2016). We retained invariant sites and SNPs (i.e. excluding complex variants) in exonic regions from the VCF of Novikova *et al.* (2016) and generated an alignment of 2,270 nucleotides, of which 653 were polymorphic. Heterozygous sites were scored with IUPAC ambiguity codes. Maximum Likelihood tree reconstruction was performed in RAxML v8.4.2 (Stamatakis, 2014) starting from 100 random parsimony trees and branch support assessed with 2,000 bootstrap replicates. *Capsella rubella* was chosen to root the phylogeny.

Data comparisons. Congruence among methods was assessed by comparing the list of loci containing either F_{ST} or CMH SNP outliers, with genes that either overlap with islands of F_{ST} and D_a divergence or with soil-specific cacti, using Venn diagrams from the R package *limma* (Ritchie et al., 2015). Only genes for which at least 500 bp could be mapped in the pooled

sequence data were considered. For each species, we finally calculated gene-wise F_{ST}, D_a, π and Tajima's D estimates for subsequent comparisons against empirical, genome-wide distributions. To assess the mechanisms that may underlie the origin of soil-specific alleles, gene-wise estimates of absolute measures of divergence D_{xy} were also included (Cruickshank & Hahn, 2014; Guerrero & Hahn, 2017).

Results

Evidence for edaphic adaptation in *A. lyrata*. Analyses of rock and soil samples from each study site (Fig. 1a, Supporting Information Table S1) indicated that four populations occur on shallow, slightly alkaline, calcium- and magnesium-rich substrates (Rendzic Leptosols) resulting from the weathering of calcareous rocks, while three populations grow on shallow, acidic, aluminium and silicon-rich substrates (Umbric Leptosols) derived from the weathering of siliceous rocks (Fig. 1b). The reciprocal germination of *A. lyrata* seeds from calcareous and siliceous populations on either soil mixture evidenced significant interaction effects of seed origin and soil treatment on time to germination (P = 0.003) but not on germination rate (P = 0.064; Fig. 1c, Supporting Information Table S1). Seeds originating from siliceous populations germinated significantly earlier (μ_{SIL} = 6.4 days, μ_{CAL} = 7.2 days, $P < 0.01$) and at a higher rate (μ_{SIL} = 84% days, μ_{CAL} = 67%, $P < 0.01$) on siliceous than on calcareous soil, indicating edaphic differentiation in siliceous populations of *A. lyrata* to siliceous soil. In contrast, *A. lyrata* originating from calcareous soils did not germinate faster (μ_{SIL} = 6.9 days, μ_{CAL} = 6.9 days) or at a higher rate (μ_{SIL} = 62%, μ_{CAL} = 69%) on calcareous soil than populations from siliceous soil, suggesting an absence of edaphic adaptation in calcareous populations at the germination stage.

Signature of divergent selection among *A. lyrata* soil ecotypes. For each of the seven *A. lyrata* populations, an average of 30 Gb of 100 bp paired-end Illumina reads were obtained by sequencing pools of genomic DNA from 30 individuals. After stringent quality filtering, about 55% of the reads were retained for subsequent analyses, for a total of 63,022,121 nucleotide positions and a total of 3,789,788 SNPs (Supporting Information Tables S2-S4). Nucleotide diversity (π) varied along the chromosomes (Supporting Information Fig. S3), but genetic differentiation (F_{ST}) between populations was overall sufficiently low (pairwise comparisons, mean F_{ST} = 0.120; Fig. 2a), to detect putative adaptive loci characterised by high divergence. Although the empirical distribution of gene-wise Tajima's D estimates indicated deviations from a neutral equilibrium model without population structure (with sample means ranging from μ = -0.26 in PER to μ = 0.17 in STR; Supporting Information Fig. S4), the consequent increase in genetic drift effects on selection analyses was minimised through (i) conservative filtering steps (*e.g.* by only retaining variants that diverged across all calcareous-siliceous comparisons) and (ii) complementary analyses within the framework of genome-wide data (*e.g.* by considering both variations in nucleotide diversity and divergence in window-based assessments). Noteworthy, signals of population structure revealed by phylogenetic analyses are not confounded with putative signals of edaphic divergence (Fig. 2c).

A total of 859 outlier SNPs was recovered with the 95% F_{ST} cut-off approach ($F_{ST\ cutoff}$ $\geq$ 0.25), versus 2,131 using CMH tests (Supporting Information Table S3). These outliers were proportionally more frequent in genic regions than expected ($P \leq 0.001$, Fisher exact test). Overall, 699 SNPs in 106 genes were shared among the two approaches and gene ontology (GO) enrichment analyses indicated that genes involved in phosphate ion transport,

cellular response to phosphate starvation, endomembrane system organisation and, to a lesser extent, cellular calcium ion homeostasis, were over-represented among outlier SNPs (Supporting Information Tables S5-S6). Altogether, these results indicate that genes involved in ion homeostasis play a central role in edaphic adaptation and ecotype divergence in *A. lyrata*. The highest number of outlier SNPs (153 CMH outliers) was found in the vacuolar phosphate transporter *VPT1* (homolog AL2G12590 of AT1G63010; Fig. 2b, Supporting Information Table S7).

We identified 346 (out of 37,487) windows (mean length ± SE: 5,230 bp ± 185 bp) with both significantly elevated F_{ST} and D_a estimates. These contained 364 genes, 48 of which had already been identified with both SNP-based approaches (Supporting Information Table S7). The cacti analyses uncovered 32 genomic regions, in which populations were phylogenetically grouped according to soil type. These genomic regions correspond to 1.1% of the genome (2,102,789 bp) and cover 233 genes (Supporting Information Table S7). Across all four inference methods, 26 genes (out of 23,272 mapped genes) showed consistent signatures of genetic divergence among edaphic ecotypes (Fig. 2d, Supporting Information Fig. S5 and Table S8). Four of these loci exhibited gene-wise D_{xy} estimates higher than the 99% quantile of the empirical distribution ($D_{xy\ cutoff} \geq 0.029$), three of which (AL2G12590, AL6G42430 and AL6G21820) had D_{xy} estimates at least twice the average within-deme π (Supporting Information Tables S7-S8 and Fig. S6).

Evidence of soil-specific copy-number variants in *A. lyrata*. Given an average of 12 Gb sequence data per individual, we identified 40,956 DEL and 1,418 TD relative to the *A. lyrata* reference genome. DEL and TD were evenly distributed across the genome (Table 1, Supporting Information Tables S9-S11). Their size distributions were skewed towards small sizes with TD being on average larger (median: 4,609 bp) than DEL (median: 348 bp; $P < 0.001$, Mann-Whitney U-test). As opposed to DEL, TD were concentrated in intragenic regions (Table 1).

We discovered both 1,623 DEL and 65 TD with considerable frequency differences between soil types (absolute frequency differences above 0.3, Fig. 3, Supporting Information Tables S10-S11). For example, a 468 bp DEL affecting the intronic region of sulphate transporter *SULTR1;1* was fixed in calcareous and absent in siliceous populations, respectively (Fig. 3a, ID: 95F_DELLY_Deletion_00029205 in Supporting Information Table S10). The same locus was identified as one of the main candidates for edaphic adaptation in *A. lyrata* in our nucleotide-based analyses (Fig. 2b, Supporting Information Table S8). Likewise, a 5,322 bp TD that causes the formation of a novel gene by fusion of two copies of glutamate receptor *GLR2.5* (paralogs AL6G21820 and AL6G21830 of AT5G11210) was close to fixation in calcareous but rare in siliceous populations, respectively (Fig. 3b, ID: 112K_PINDEL_Tandem_Duplication_1089 in Supporting Information Table S11). We also detected soil-specific DEL and TD in genes devoid of any obvious selection signals in the nucleotide dataset, such as a 469 bp DEL affecting the intronic region of flowering regulator *MAF3* (homolog AL233U10010 of AT5G65060; 170K_PINDEL_Deletion_231954 in Supporting Information Table S10), or a 7,518 bp TD affecting two paralogs of the germination-specific cysteine protease *CP1* (homologs AL7G13630 and AL7G13640 of AT4G36880; 88A_PINDEL_Tandem_Duplication_223 in Supporting Information Table S11).

Evidence of shared polymorphisms between *A. lyrata* and *A. arenosa*. As for *A. lyrata*, an average of 30 Gb of 100 bp paired-end Illumina reads were obtained for each population of *A.*

arenosa (Supporting Information Table S2). After quality filtering and PCR duplicate removal, about 41% of the reads were retained for subsequent analyses, for a total of 63,294,736 nucleotide positions (Supporting Information Table S2 and S12). Similarly, π varied along the chromosomes (Supporting Information Fig. S7), and mean pairwise F_{ST} was overall very low (mean F_{ST} = 0.072; Supporting Information Fig. S7). Since the empirical distribution of gene-wise Tajima's *D* estimates rejected the neutral equilibrium model without population structure for all but one population (with sample means ranging from μ = -0.74 in HAR to μ = 0.00 in DHL; Supporting Information Fig. S7) and phylogenetic analyses indicated a consistent genetic divergence between low-altitudinal, calcareous populations and high-altitudinal, siliceous taxa (data not shown), the dataset of *A. arenosa* was only meant for comparison with *A. lyrata*.

Given a 95% F_{ST} cut-off of 0.17, we detected 6,294,639 SNPs in *A. arenosa*, of which 44,115 and 25,508 were F_{ST} and CMH outliers, respectively (Supporting Information Fig. S7-S8 and Tables S12-S13). Genes involved in ion transport, response to nitrate and cellular ion homeostasis were over-represented among outlier SNPs (Supporting Information Tables S14-S15), indicating that genetic differentiation between soil ecotypes of *A. arenosa* targets similar biochemical pathways as those identified in the sister species *A. lyrata* (Fig. 4a). The higher number of SNP outliers detected in *A. arenosa* in respect to *A. lyrata* likely results from confounding effects of genetic drift and altitudinal divergence in the former species.

Altogether, nine genes showed consistent signals of divergent selection in their coding regions among soil ecotypes in both *A. lyrata* and *A. arenosa* (Table 2, Supporting Information Fig. S9, Table S16). Two of these loci exhibited gene-wise D_{xy} estimates higher than the 99% quantile of the empirical distribution ($D_{xy\ cutoff} \geq 0.038$), one of which (AL6G42430) had D_{xy} estimates at least twice higher than the average within-deme π (Supporting Information Table S16 and Fig. S10). The phylogenetic analysis of both *A. lyrata* and *A. arenosa* consensus sequences further highlighted 13,065 bp (0.07‰ of the genome) encompassing two of these nine candidate genes (homolog AL6G42430 encoding *SULTR1;1* and homolog AL5G25250), which entailed polymorphisms that grouped taxa as a function of soil type rather than species identity (Fig. 4b, Table 2, Supporting Information Table S17). The high-affinity sulphate transporter *SULTR1;1* alone contains 17 SNPs that are shared across soil ecotypes between the two *Arabidopsis* species (Fig. 4c). The phylogenetic tree inferred from 72 *Arabidopsis* accessions shows that accessions from calcareous and siliceous populations fall in separate clades along with other species of the genus. This suggests that soil-specific *SULTR1;1* alleles evolved prior to the divergence between *A. lyrata* and *A. arenosa* (Fig. 5).

Discussion

Differential response of soil ecotypes to the respective substrates during early life stage. *Arabidopsis lyrata* germinates both on calcareous and siliceous soils, but the reciprocal germination experiment reveals significant GxE effects, in agreement with similar studies on calcicole-calcifuge sister lineages (Anderson, 1982; Ramakrishnan, 1965). Siliceous populations indeed exhibit a faster and higher germination rate on acidic than alkaline soils, but no differential response was detected in calcareous populations. Although this experiment cannot assess the effect of soil type on long-term survival, seed germination represents a sensitive early phase in the life cycle that may contribute to ecological differentiation between soil ecotypes (Baskin & Baskin, 2014; Fenner & Thompson, 2005).

Seedlings usually prefer colonising open habitats devoid or depleted of competitors (Bullock, 2000). Such 'vegetation gaps' differ environmentally from the surrounding, intact areas with regard to light intensity, temperature fluctuation and nutrient supply (the main determinants of seed germination). By detecting gap-signals, seedlings may increase their survival rate by germinating when resource competition is reduced. Experiments with *Plantago lanceolata* for example showed that seeds buried in bare soil germinated more readily than in vegetated patches, because seeds detected more nitrate in the former than in the latter habitat (so called gap-detection mechanism; Pons, 1989).

Consistent signatures of genetic divergence at loci involved in ion homeostasis. Calcareous and siliceous soils substantially differ with respect to their chemical and physical properties (Bothe, 2015; Kinzel, 1983; Lee, 1999). As expected, genomic comparisons of *A. lyrata* soil ecotypes reveal an enrichment of genes involved in ion transport among candidates of divergent selection, such as glutamate receptors *GLR2.5* (Spalding & Harper, 2011) and potassium uptake permeases *KUP9/KUP10* (homologs AL1G45650 and AL7G33950 of AT1G31120 and AT4G19960, respectively; Ahn, Shin, & Scachtman, 2004), which jointly regulate cellular calcium and potassium concentrations.

The detection of multiple outlier SNPs in genes involved in phosphate homeostasis, such as the vacuolar phosphate transporter *VPT1* (Liu et al., 2015), is consistent with differential response to soil chemistry. Apatite, for example, constitutes the primary source of inorganic phosphorus in rocks and is more soluble at low than at neutral-to-alkaline pH (Bothe, 2015; D. L. Jones & Oburger, 2011; Kinzel, 1983; Lee, 1999). Biotic factors may further drive edaphic adaptation in *A. lyrata*, owing to chemically-driven changes in plant-soil microbial interactions, hence solubilisation mechanisms in the rhizosphere (D. L. Jones & Oburger, 2011). Interestingly, our results do not provide direct evidence for adaptation to physical factors, despite the general expectation that calcareous populations face increased drought and temperature stress, due to the low water retention capacity of calcareous rocks (Bothe, 2015; Kinzel, 1983).

The presence of SNP outliers in several genes involved in seed germination, such as the abscisic acid (ABA) catabolic gene *CYP707A2* (homolog AL4G23660 of AT2G29090; Matakiadis et al., 2009), suggests a genetic basis of the observed shift in seed dormancy between calcareous and siliceous populations of *A. lyrata* in reciprocal germination experiments. Environmental cues such as temperature, light, and nutrient supply individually or interactively influence seed dormancy through downstream effects on abscisic acid (ABA) and gibberellin (GA) content. ABA and GA antagonistically regulate seed germination, the former maintaining and the latter releasing seed dormancy, respectively. Specifically, *CYP707A2*, which encodes an ABA catalytic enzyme, causes a decrease in ABA seed content in *A. thaliana* (thereby releasing dormancy) after induction by nitrate or nitric oxide (Matakiadis et al., 2009; Yan et al., 2016).

Aside from its role as nutrient, nitrate has long been recognised as a signalling molecule in seed germination (reviewed in Duermeyer et al., 2018), along with other forms of nitrogen such as nitric oxide and ammonium (Bethke, Libourel, & Jones, 2007; Hendricks & Taylorson, 1974). For example, the amount of nitrate applied to mother plants or imbibed seeds negatively correlates with the magnitude of seed dormancy in *Arabidopsis* (Alboresi et al., 2005). How different nitrogen compounds interact to mediate seed germination is not fully resolved though (Duermeyer et al., 2018). Given that light and temperature were kept constant in our germination trials, we hypothesise that the availability of nitrogen as ammonium rather than nitrate in siliceous soils (Bothe, 2015; Kinzel, 1983; Lee, 1999)

mediated differential responses in seed dormancy in *A. lyrata* soil ecotypes. The detection of multiple outlier SNPs in the high-affinity nitrate transporter *NRT2.1* (homolog AL1G18450 of AT1G08090; Remans et al., 2006) further corroborates our assumption. A possible effect of different soil microbial communities on seed germination remains to be tested, because unsterilised soil was used for our experiment.

Complementary role of CNVs during edaphic adaptation. In line with growing evidence that CNVs contribute to edaphic adaptation in plants (Hanikenne et al., 2013; Hanikenne et al., 2008; Maron et al., 2013; Suryawanshi et al., 2016), we found soil-specific DEL and TD in genic regions of *A. lyrata*. Of the 26 genes exhibiting the most consistent patterns of genetic differentiation among soil ecotypes at the nucleotide level, the high-affinity sulphate transporter *SULTR1;1* (Rouached, Secco, & Arpat, 2009) and the glutamate receptors *GLR2.5* (Spalding & Harper, 2011) also entail soil-specific CNVs.

Whether the intronic DEL in *SULTR1;1* causes regulatory changes or alternatively constitutes a hitchhiking neutral polymorphism affected by linked selection remains to be investigated. Similarly, the genomic data do not allow us to ascertain that the TD spanning the two *GLR2.5* paralogs is adaptive, but gene duplications were repeatedly shown to affect transcript expression levels and to favour neo-functionalisation during evolution (Long, Betrán, Thornton, & Wang, 2003; Zhang, Gu, Hurles, & Lupski, 2009). The detection of soil-specific CNVs in genes devoid of obvious nucleotide-based selection signals, such as the intronic DEL in flowering time regulator *MAF3* (Suter, Rüegg, Zemp, Hennig, & Widmer, 2014) and the TD spanning two paralogs of the seed dormancy regulator *CP1* (Piskurewicz et al., 2016), suggest that CNVs contribute to genetic variation underlying edaphic adaptation. It is possible that outlier SNPs were missed in some of these regions, because CNVs may potentially affect read mapping and associated SNP identification.

Shared adaptive genetic variation between sister species. Sequences that diverged between edaphic ecotypes of both *A. lyrata* and *A. arenosa* are enriched with highly differentiated outlier SNPs located in genes involved in ion transport. Nine of these loci further exhibit consistent signatures of divergent selection across both species. In two of these genes, the two species also share the same SNPs across soil ecotypes, overall strengthening our inference of adaptive genomic divergence in response to edaphic differences in both species. Comparisons with studies on adaptation to serpentine soils further reveal overlapping targets of selection. For example, genomic analysis of tetraploid *A. arenosa* subsp. *arenosa* (L.) Lawalrée populations uncovered strong signals of edaphic adaptation in *SULTR1;1* and *KUP9* (Arnold et al., 2016). Similarly, Turner and collaborators (2010; 2008) also detected strong differentiation at the boric acid channel *NIP5;1* (homolog AL6G44270 of AT4G10380; Kato, Miwa, Takano, Wada, & Fujiwara, 2009) between North American populations of *A. lyrata* subsp. *lyrata* O'Kane & Al-Shehbaz.

Several factors may contribute to recurrent patterns of genetic divergence over the course of evolution. Strong and persistent selective pressures on sulphur availability for glucosinolate (GS) biosynthesis (hence defence mechanisms; Dubuis, Marazzi, Städler, & Mauch, 2005) may be at the origin of ancient allelic variation in *SULTR1;1* that differs between calcareous and siliceous populations (Falk, Tokuhisa, & Gershenzon, 2007; Hopkins, van Dam, & van Loon, 2009). Adaptive constraints may then lead to the repeated recruitment of pre-existing alleles from the pool of standing genetic variation, as shown for the *EDA* locus in threespine sticklebacks (Colosimo et al., 2005; F. C. Jones et al., 2012) or genes involved in the aryl hydrocarbon receptor (*AHR*) signalling pathway in killifish (Reid et al., 2016).

Alternatively, adaptive alleles may originate from repeated genetic exchanges between lineages, as recently demonstrated in the genus *Arabidopsis*, as a result of recurrent range shifts during Pleistocene glaciation and deglaciation cycles (Arnold et al., 2016; Hohmann & Koch, 2017; Novikova et al., 2016; Novikova, Hohmann, & Van de Peer, 2018; Schmickl & Koch, 2011).

The high absolute divergence (estimated by D_{xy}) between alleles and demes, but low within-deme diversity (estimated by π) at *SULTR1;1* indicates recurrent selection on an ancient polymorphism on each soil type (Cruickshank & Hahn, 2014; Guerrero & Hahn, 2017). Having ten accessions of *A. thaliana* (L.) Heynh. included in the gene tree indeed suggests that soil-specific *SULTR1;1* alleles evolved prior to the divergence of any perennial *Arabidopsis* and that its origin might even predate the diversification of the genus. The lack of shared identity-by-descent (IBD) blocks across *SULTR1;1* between *A. lyrata* and *A. arenosa* (Novikova et al., 2016 and pers. comm. to A.G.) and the presence of species-specific SNPs in this locus leads us to favour a scenario of repeated selective sweeps on standing genetic variation over the hypothesis of recent introgression between these two sister species as a source of the divergent alleles.

The retention of two soil-specific *SULTR1;1* alleles in *A. lyrata* and *A. arenosa*, which exhibit wide distribution ranges and large effective population sizes, contrasts with the single, calcareous- or siliceous-specific allele detected in endemic *Arabidopsis* taxa with limited distribution ranges and narrow ecological niches (Fig. 5; Clauss & Koch, 2006; Koch et al., 2008). For example, *A. croatica* (Schott ex Nyman & Kotschy) O'Kane & Al-Shehbaz is restricted to calcareous bedrocks in Croatian mountains, *A. cebennensis* (DC.) O'Kane & Al-Shehbaz solely occurs on calcareous or amphibolitic bedrock types in the Massif Central (France), and *A. pedemontana* (Boiss.) O'Kane & Al-Shehbaz is exclusively found in Piedmont (Italy) on siliceous bedrock types such as granite, diorite and gneiss. It is further remarkable that the only annual species, *A. thaliana*, with no evident soil preferences, does not carry any of these alleles (Hoffmann, 2002).

This study integrates experimental and genomic evidence, which together support the notion that physiological adaptation to soil chemistry underlies adaptation to calcareous and siliceous soils in *A. lyrata* and its sister species *A. arenosa*. Our results suggest that edaphic adaptation is mediated by selection acting on early life-stages, and is targeting genes involved in diverse functions, including ion uptake, homeostasis and cellular starvation response. High nucleotide diversity in these self-incompatible species has provided the foundation for selective sweeps at single-nucleotide and copy-number variants associated with adaptation to soil types. Some of the candidate genes further show consistent sweep signatures across species boundaries, making them suitable for functional studies aimed at validating their presumed adaptive roles. Altogether, this work contributes towards our understanding of how subtle edaphic changes shape patterns of adaptive differentiation in plants.

Acknowledgements

This work was supported by the German Research Foundation (DFG SPP 1529 to M.A.K. and A.W). We are indebted to F. Kolář, Z. Kyselová, G. Muir, L. Paule and R. Schmickl for help with collections. We thank C. Michel for assistance in the lab, F. Berger-Glarner for managing the seed collection, M. Frei and E. Zürcher for taking care of the plants, P. Novikova for genome alignments and IBD analyses, K. Barmettler and R. Smith for soil and rock analyses, and all members of the Plant Ecological Genetics group, in particular M. Paris

and T. Städler, for fruitful discussions. We further acknowledge three anonymous reviewers for providing helpful comments on earlier versions of this manuscript. Data produced and analysed in this paper were generated in collaboration with S. Zoller and N. Zemp from the ETH Zurich Genetic Diversity Centre (GDC).

References

Abyzov, A., Urban, A. E., Snyder, M., & Gerstein, M. (2011). CNVator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Research, 21*(6), 974-984.

Ahn, S. J., Shin, R., & Scachtman, D. P. (2004). Expression of *KT/KUP* genes in *Arabidopsis* and the role of root hairs in K^+ uptake. *Plant Physiology, 134*, 1135-1145.

Akey, J. M., Zhang, G., Zhang, K., Jin, L., & Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Research, 12*, 1805-1814.

Alboresi, A., Gestin, C., Leydecker, M.-T., Bedu, M., Meyer, C., & Truong, H.-N. (2005). Nitrate, a signal relieving seed dormancy in *Arabidopsis*. *Plant, Cell and Environment, 28*, 500-512.

Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews in Genetics, 12*, 363-376.

Anderson, C. A. (1982). The effect of calcium on the germination, growth and mineral nutrition of acidic and calcareous populations of *Eucalyptus obliqua* L'Hérit. *Plant and Soil, 69*, 213-233.

Arnold, B. J., Lahner, B., DaCosta, J. M., Weisman, C. M., Hollister, J. D., Salt, D. E., . . . Yant, L. (2016). Borrowed alleles and convergence in serpentine adaptation. *Proceedings of the National Academy of Science, USA, 113*(29), 8320-8325.

Bamberg, S. A., & Major, J. (1968). Ecology of the vegetation and soils associated with calcareous parent materials in three alpine regions of Montana. *Ecological Monographs, 38*(2), 127-167.

Baskin, C. C., & Baskin, J. M. (2014). *Seeds: Ecology, Biogeography, and Evolution of Dormancy and Germination*. Amsterdam, The Netherlands: Elsevier Academic Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4*: Linear mixed-effects models using Eigen and S4. R package version 1.1-7: http://cran.r-project.org/package=lme4.

Beadle, N. C. W. (1953). The edaphic factor in plant ecology with a special note on soil phosphates. *Ecology, 34*(2), 426-428.

Beissinger, T. M., Rosa, G. J. M., Kaeppler, S. M., Gianola, D., & de Leon, N. (2015). Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genetics, Selection, Evolution, 47*, 30.

Bellos, E., Johnson, M. R., & Coin, L. J. M. (2012). cnvHiTSeq: integrative models for high-resolution copy number variation detection and genotyping using population sequencing data. *Genome Biology, 13*, R120.

Bethke, P. C., Libourel, I. G. L., & Jones, R. L. (2007). Nitric oxide in seed dormancy and germination. In K. J. Bradford & H. Nonogaki (Eds.), *Seed Development, Dormancy and Germination*. New Delhi, India: Blackwell Publishing Ltd.

Blume, H.-P., Brümmer, G. W., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretzschmar, R., . . . Wilke, B.-M. (2010). *Lehrbuch der Bodenkunde*. Heidelberg, Germany: Springer Verlag.

Bothe, H. (2015). The lime-silicate question. *Soil Biology & Biochemistry, 89*, 172-183.

Braun-Blanquet, J. (1964). *Pflanzensoziologie: Grundzüge der Vegetationskunde*. Wien, Austria: Springer-Verlag.
Bullock, J. M. (2000). Gaps and seedling colonization. In M. Fenner & K. Thompson (Eds.), *The Ecology of Seeds* (pp. 375-395). Cambridge, UK: Cambridge University Press.
Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution, 15*(5), 538-543.
Clauss, M. J., & Koch, M. A. (2006). Poorly known relatives of *Arabidopsis thaliana*. *Trends in Plant Science, 11*(9), 449-459.
Clauss, M. J., & Mitchell-Olds, T. (2006). Population genetic structure of *Arabidopsis lyrata* in Europe. *Molecular Ecology, 15*, 2753-2766.
Clymo, R. S. (1962). An experimental approach to part of the calcicole problem. *Journal of Ecology, 50*(3), 707-731.
Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal Jr., G., Dickson, M., Grimwood, J., . . . Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science, 307*, 1928-1933.
Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology, 23*, 3133-3157.
Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews in Genetics, 14*, 262-274.
Dassanayake, M., Oh, D.-H., Haas, J. S., Hernandez, A., Hong, H., Ali, S., . . . Cheeseman, J. M. (2011). The genome of the extremophile crucifer *Thellungiella parvula*. *Nature Genetics, 43*(9), 913-920.
Davies, M. S., & Snaydon, R. W. (1973a). Physiological differences among populations of *Anthoxanthum odoratum* L. collected from the Park Grass Experiment, Rothamsted. I. Response to calcium *Journal of Applied Ecology, 10*(1), 33-45.
Davies, M. S., & Snaydon, R. W. (1973b). Physiological differences among populations of *Anthoxanthum odoratum* L. collected from the Park Grass experiment, Rothamsted. II. Response to aluminium. *Journal of Applied Ecology, 10*(1), 47-55.
de Mita, S., Thuilet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology, 22*(5), 1383-1399.
Dubuis, P.-H., Marazzi, C., Städler, E., & Mauch, F. (2005). Sulphur deficiency causes a reduction in antimicrobial potential and leads to increased diseases susceptibility of oilseed rape. *Journal of Phytopathology, 153*, 27-36.
Duermeyer, L., Khodapanahi, E., Yan, D., Krapp, A., Rothstein, S. J., & Nambara, E. (2018). Regulation of seed dormancy and germination by nitrate. *Seed Science Research*. doi:10.1017/S096025851800020X
Ellenberg, H. (1958). Bodenreaktion (einschliesslich Kalkfrage). In W. Ruhland (Ed.), *Handbuch der Pflanzenphysiologie* (Vol. 4, pp. 638–708). Berlin, Germany: Springer-Verlag.
Eshel, G., Shaked, R., Kazachkova, Y., Khan, A., Eppel, A., Cisneros, A., . . . Barak, S. (2017). *Anastatica hierochuntica*, an *Arabidopsis* desert relative, is tolerant to multiple abiotic stresses and exhibits species-specific and common stress tolerance strategies with its halopyhtic relative, *Eutrema* (*Thellungiella*) *salsugineum*. *Frontiers in Plant Science, 7*, 1992.

Falk, K. L., Tokuhisa, J. G., & Gershenzon, J. (2007). The effect of sulfur nutrition on plant glucosinolate content: phyiology and molecular mechanisms. *Plant Biology, 9*, 573-581.
Fenner, M., & Thompson, K. (2005). *The Ecology of Seeds*. Cambridge, UK: Cambridge University Press.
Ferretti, L., Ramos-Onsins, S. E., & Pérez-Enciso, M. (2013). Population genomics from pool sequencing. *Molecular Ecology, 22*, 5561-5576.
Flood, P. J., & Hancock, A. M. (2017). The genomic bais of adaptation in plants. *Current Opinion in Plant Biology, 36*, 88-94.
Futschik, A., & Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics, 186*, 207-218.
Gigon, A. (1971). *Vergleich alpiner Rasen auf Silikat- und auf Karbonatboden: Konkurrenz- und Stickstofformenversuche sowie standortskundliche Untersuchungen im Nardetum und im Seslerietum bei Davos.* (48 PhD), ETH Zurich, Zurich, Switzerland.
Gigon, A. (1987). A hierarchic approach in causal ecosystem analysis: The calcifuge-calcicole problem in Alpine grasslands. In E.-D. Schulze & H. Zwölfer (Eds.), *Ecological Studies, Potentials and Limitations of Ecosystem Analysis* (Vol. 61, pp. 228-244). Berlin, Germany: Springer-Verlag.
Givnish, T. J. (2010). Ecology of plant speciation. *Taxon, 59*(5), 1326-1366.
Gould, B., McCouch, S., & Geber, M. (2014). Variation in soil aluminium tolerance genes is associated with loal adaptation to soils at the Park Grass Experiment. *Molecular Ecology, 23*, 6058-6072.
Gould, B., McCouch, S., & Geber, M. (2015). *De novo* transcriptome assembly and identification of gene candidates for rapid evolution of soil Al tolerance in *Anthoxanthum odoratum* at the long-term Park Grass Experiment. *PLoS One, 10*(7), e0124424.
Grime, J. P., & Curtis, A. V. (1976). The interaction of drought and mineral nutrient stress in calcareous grassland. *Journal of Ecology, 64*(3), 975-988.
Guerrero, R. F., & Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology, 2017*, 1-7.
Hanikenne, M., Kroymann, J., Trampczynska, A., Bernal, M., Motte, P., Clemens, S., & Krämer, U. (2013). Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation. *PLoS Genetics, 9*(8), e1003707.
Hanikenne, M., Talke, I. N., Haydon, M. J., Lanz, C., Nolte, A., Motte, P., . . . Krämer, U. (2008). Evolution of metal hyperaccumulation required *cis*-regulatory changes and triplication of *HMA4*. *Nature, 453*, 391-395.
Hendricks, S. B., & Taylorson, R. B. (1974). Promotion of seed germination by nitrate, nitrite, hydroxylamine, and ammonium salts. *Plant Physiology, 54*, 304-309.
Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics, 169*, 2335-2336.
Hoffmann, M. H. (2002). Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). *Journal of Biogeography, 29*, 125-134.
Hohmann, N., & Koch, M. A. (2017). An *Arabidopsis* introgression zone studied at high spatio-temporal resolution: interglacial and multiple genetic contact exemplified using whole nuclear and plastid genomes. *BMC Genomics, 18*, 810.
Hohmann, N., Schmickl, R., Chiang, T.-Y., Lučanová, M., Kolář, F., Marhold, K., & Koch, M. A. (2014). Taming the wild: resolving the gene pools of non-model *Arabidopsis* lineages. *BMC Evolutionary Biology, 14*, 224.

Hohmann, N., Wolf, E. M., Lysak, M. A., & Koch, M. A. (2015). A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *The Plant Cell, 27*(10), 2770-2784.

Hopkins, R. J., van Dam, N. M., & van Loon, J. J. A. (2009). Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annual Review of Entomology, 54*, 57-83.

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., . . . Guo, Y.-L. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics, 43*(5), 476-481.

Hughes, J. D. (1985). Theophrastus as ecologist. *Environmental Review, 9*(4), 296-306.

Jones, D. L., & Oburger, E. (2011). Solubilization of phosphorus by soil microorganisms. In E. K. Bünemann, A. Oberson, & E. Frossard (Eds.), *Phosphorus in Action* (pp. 169-198). Heidelberg, Germany: Springer Verlag.

Jones, F. C., Grabheer, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., . . . Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature, 484*, 55-61.

Kato, Y., Miwa, K., Takano, J., Wada, M., & Fujiwara, T. (2009). Highly boron deficiency-tolerant plants generated by enhanced expression of *NIP5;1*, a boric acid channel. *Plant Cell Physiology, 50*(1), 58-66.

Kinzel, H. (1983). Influence of limestone, silicates and soil pH on vegetation. In O. L. Lange, P. S. Nobel, C. B. Osmond, & H. Ziegler (Eds.), *Physiological Plant Ecology III* (pp. 201-244). Berlin-Heidelberg: Springer

Koch, M. A., Wernisch, M., & Schmickl, R. (2008). *Arabidopsis thaliana*'s wild relatives: an updated overview on systematics, taxonomy and evolution. *Taxon, 57*(3), 933-943.

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics, 27*(4), 3435-3436.

Kolář, F., Fuxová, G., Záveská, E., Nagano, A. J., Hyklová, L., Lučanová, M., . . . Marhold, K. (2016). Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*. *Molecular Ecology, 25*, 3929-3949.

Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences, 279*(1749), 5048-5057. doi:10.1098/rspb.2012.1108

Krämer, U. (2010). Metal hyperaccumulation in plants. *Annual Review of Plant Biology, 61*, 517-534.

Kruckeberg, A. R. (1986). An essay: the stimulus of unusual geologies for plant speciation. *Systematic Botany, 11*(3), 455-463.

Lee, J. A. (1999). The calcicole-calcifuge problem revisited. *Advances in Botanical Research, 29*, 1-30.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv,* , 1303.3997v1302.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Li, W., & Olivier, M. (2013). Current analysis platforms and methods for detecting copy number variation. *Physiological Genomics, 45*, 1-16.

Liu, J., Yang, L., Luan, M., Wang, Y., Zhang, C., Zhang, B., . . . Luan, S. (2015). A vacuolar phosphate transporter essential for phosphate homeostasis in *Arabidopsis*. *Proceedings of the National Academy of Science, USA, 112*(47), e6571-e6578.

Long, M., Betrán, E., Thornton, K., & Wang, W. (2003). The origin of new genes: glimpses from the young and old. *Nature Reviews in Genetics, 4*, 865-875.

Maron, L. G., Guimarães, C. T., Kirst, M., Albert, P. S., Birchler, J. A., Bradbury, P. J., . . . Kochian, L. V. (2013). Aluminium tolerance in maize is associated with higher *MATE1* gene copy number. *Proceedings of the National Academy of Science, USA, 110*(13), 5241-5246.

Mason, H. L. (1946). The edaphic factor in narrow endemism. I. The nature of environmental influences. *Madroño, 8*(7), 209-226.

Matakiadis, T., Alboresi, A., Jikumaru, Y., Tatematsu, K., Pichon, O., Renou, J.-P., . . . Truong, H.-N. (2009). The *Arabidopsis* abscisic acid catabolic gene *CYP707A2* plays a key role in nitrate control of seed dormancy. *Plant Physiology, 149*, 949-960.

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution, 28*(11), 659-669.

Michalet, R., Gandoy, C., Joud, D., & Pagès, J.-P. (2002). Plant community composition and biomass on calcareous and siliceous substrates in the northern French Alps: comparative effects of soil chemistry and water status. *Arctic, Antarctic, and Alpine Research, 34*(1), 102-113.

Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., . . . 1000 Genomes Project. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature, 470*(7332), 59-65.

Muir, G., Bergero, R., Charlesworth, D., & Filatov, D. A. (2011). Does local adaptation cause high population differentiation of *Silene latifolia* Y chromosomes? *Evolution, 65*(12), 3368-3380.

Nei, M., & Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Science, USA, 76*(10), 5269-5273.

Novikova, P. Y., Hohmann, N., Nizhynska, V., Tschchimatsu, T., Ali, J., Muir, G., . . . Nordborg, M. (2016). Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics, 48*, 1077–1082.

Novikova, P. Y., Hohmann, N., & Van de Peer, Y. (2018). Polyploid Arabidopsis species originated around recent glaciation maxima. *Current Opinion in Plant Biology, 42*, 8-15.

Piskurewicz, U., Iwasaki, M., Susaki, D., Megies, C., Kinoshita, T., & Lopez-Molina, L. (2016). Dormancy-specific imprinting underlies maternal inheritance of seed dormancy in *Arabidopsis thaliana*. *eLife, 5*, e19573.

Pons, T. L. (1989). Breaking of seed dormancy by nitrate as a gap detection mechanism. *Annals of Botany, 63*, 139-143.

R Development Core Team. (2018). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rajakaruna, N. (2004). The edaphic factor in the origin of plant species. *International Geology Review, 46*, 471-478.

Rajakaruna, N. (2017). Lessons on evolution from the study of edaphic specialization. *Botanical Review*.

Rajakaruna, N., & Boyd, R. S. (2008). Edaphic factor. In S. E. Jorgensen & B. Fath (Eds.), *Encyclopedia of Ecology* (1st ed., pp. 1201-1207). Oxford, UK: Elsevier.

Ramakrishnan, P. S. (1965). Edaphic ecotypes in *Euphorbia thymifolia* L.: I. Seed germination. *Journal of Ecology, 53*(1), 157-162.

Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., & Korbel, J. O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics, 28*(18), i333-i339.

Rawat, V., Abdelsamad, A., Pietzenuk, B., Seymour, D. K., Koenig, D., Weigel, D., . . . Schneeberger, K. (2015). Improving the annotation of *Arabidopsis lyrata* using RNA-Seq data. *PLoS One, 10*(9), e0137391.

Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., . . . Whitehead, A. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science, 354*, 1305-1308.

Rellstab, C., Zoller, S., Tedder, A., Gugerli, F., & Fischer, M. C. (2013). Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. *PLoS One, 8*(11), e80422.

Remans, T., Nacry, P., Pervent, M., Girin, T., Tillard, P., Lepetit, M., & Gojon, A. (2006). A central role for the nitrate transporter NRT2.1 in the integrated morphological and physiological responses of the root system to nitrogen limiation in *Arabidopsis*. *Plant Physiology, 140*, 909-921.

Ritchie, M. E., Phipson, B.-., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research, 43*(7), e47.

Rouached, H., Secco, D., & Arpat, A. B. (2009). Getting the most sulfate from soil: regulation of sulfate uptake transporters in *Arabidopsis*. *Journal of Plant Physiology, 166*, 893-900.

Ryan, P. R., Tyerman, S. D., Sasaki, T., Furuichi, T., Yamamoto, Y., Zhang, W. H., & Delhaize, E. (2011). The identification of aluminium-resistance genes provides opportunities for enhancing crop production on acid soils. *Journal of Experimental Botany, 62*(1), 9-20.

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

Schmickl, R., & Koch, M. A. (2011). *Arabidopsis* hybrid speciation processes. *Proceedings of the National Academy of Science, USA, 108*(34), 14192–14197.

Schmickl, R., Paule, J., Klein, J., Marhold, K., & Koch, M. A. (2012). The evolutionary history of the *Arabidopsis arenosa* complex: diverse tetraploids mask the Western Carpathian center of species and genetic diversity. *PLoS One, 7*(8), e42691.

Smeds, L., & Künstner, A. (2011). ConDeTri - a content dependent read trimmer for illumina data. *PLoS One, 6*(10), e26314.

Spalding, E. P., & Harper, J. F. (2011). The ins and outs of cellular Ca^{2+} transport. *Current Opinion in Plant Biology, 14*, 715-720.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313.

Supek, F., Bosňjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One, 6*(7), e21800.

Suryawanshi, V., Talke, I. N., Weber, M., Eils, R., Brors, B., Clemens, S., & Krämer, U. (2016). Between-species differences in gene copy number are enriched among functions critical for adaptive evolution in *Arabidopsis halleri*. *BMC Genomics, 17*(S13), 1034.

Suter, L., Rüegg, M., Zemp, N., Hennig, L., & Widmer, A. (2014). Gene regulatory variation mediates flowering responses to vernalization along an altitudinal gradient in *Arabidopsis thaliana*. *Plant Physiology*.

Szkiba, D., Kapun, M., von Haeseler, A., & Gallach, M. (2014). SNP2GO: functional analysis of genome-wide association studies. *Genetics, 197*, 285-289.

Tansley, A. G. (1917). On competition between *Galium saxatile* L. (*G. hercynium* Weig.) and *Galium sylvestre* Poll. (*G. asperum* Schreb.) on different types of soil. *Journal of Ecology, 5*, 173-179.

Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T., & Nuzhdin, S. V. (2010). Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics, 42*(3), 260-263.

Turner, T. L., Von Wettberg, E. J., & Nuzhdin, S. V. (2008). Genomic analysis of differentiation between soil types reveals candidate genes for local adaptation in *Arabidopsis lyrata*. *PLoS One, 3*(9), e3183.

Unger, F. (1836). *Über den Einfluss des Bodens auf die Verteilung der Gewächse, nachgewiesen in der Vegetation des nordöstlichen Tirols*. Wien, Austria: Rohrmann & Schweigerd.

Waser, N. M., & Campbell, D. R. (2004). Ecological speciation in flowering plants. In U. Dieckmann, M. Doebeli, J. A. J. Metz, & D. Tautz (Eds.), *Adaptive Speciation* (pp. 264-277). Cambridge, UK: Cambridge University Press.

Wright, K. M., Hellsten, U., Xu, C., Jeong, A. L., Sreedasyam, A., Chapman, J. A., . . . Willis, J. H. (2015). Adaptation to heavy-metal contaminated environments proceeds via selection on pre-existing genetic variation. *bioRxiv*. doi:10.1101/029900

Wu, H.-J., Zhang, Z., Wang, J.-Y., Oh, D.-H., Dassanayake, M., Liu, B., . . . Xie, Q. (2012). Insights into salt tolerance from the genome of *Thellungiella salsuginea*. *Proceedings of the National Academy of Science, USA, 109*(30), 12219-12224.

Xi, R., Lee, S., & Park, P. J. (2012). A survey of copy-number variation detection tools based on high-throughput sequencin data. *Current Protocols in Human Genetics, S74*, Unit7.19.

Yan, D., Easwaran, V., Chau, V., Okamoto, M., Ierullo, M., Kimura, M., . . . Nambara, E. (2016). NIN-like protein 8 is a master regulator of nitrate-promoted seed germination in *Arabidopsis*. *Nature Communications, 7*, 13179.

Ye, K., Schulz, M. H., Long, Q., Apweiler, R., & Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics, 25*(21), 2865-2871.

Zamani, N., Russell, P., Lantz, H., Hoeppner, M. P., Meadows, J. R. S., Vijay, N., . . . Grabherr, M. (2013). Unsupervised genome-wide recognition of local relationship patterns. *BMC Genomics, 14*, 347.

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. *Annual Review of Genomics and Human Genetics, 10*, 451-481.

Zichner, T., Garfield, D. A., Rausch, T., Stütz, A. M., Cannavó, E., Braun, M., . . . Korbel, J. O. (2013). Impact of genomic structural variation in *Drosophila melanogaster* based on population-scale sequencing. *Genome Research, 23*(3), 568-579.

Żmieńko, A., Samelak, A., Kozłowski, P., & Figlerowicz, M. (2014). Copy number polymorphism in plant genomes. *Theoretical and Applied Genetics, 127*, 1-18.

Data Accessibility Statement

Raw Illumina sequence data are available in NCBI's Sequence Read Archive (SRA) under BioProject PRJNA493227 (BioSamples SAMN10132500-SAMN10132528 and SAMN04432147-SAMN04432152; see Supporting Information Table S2 and S9). Further data that support the findings of this study are available from the corresponding author on reasonable request.

Author Contributions

A.G., M.C.F., S.F., M.R., R.K., M.A.K. and A.W. designed the experiments and formulated predictions; A.G. and G.M. contributed plant collections; L.S. performed and analysed the germination experiment; X.L. analysed the copy-number variants; S.F. performed the phylogenetic analyses; A.G. analysed the pooled sequence data and drafted the manuscript; all authors contributed to editing the manuscript.

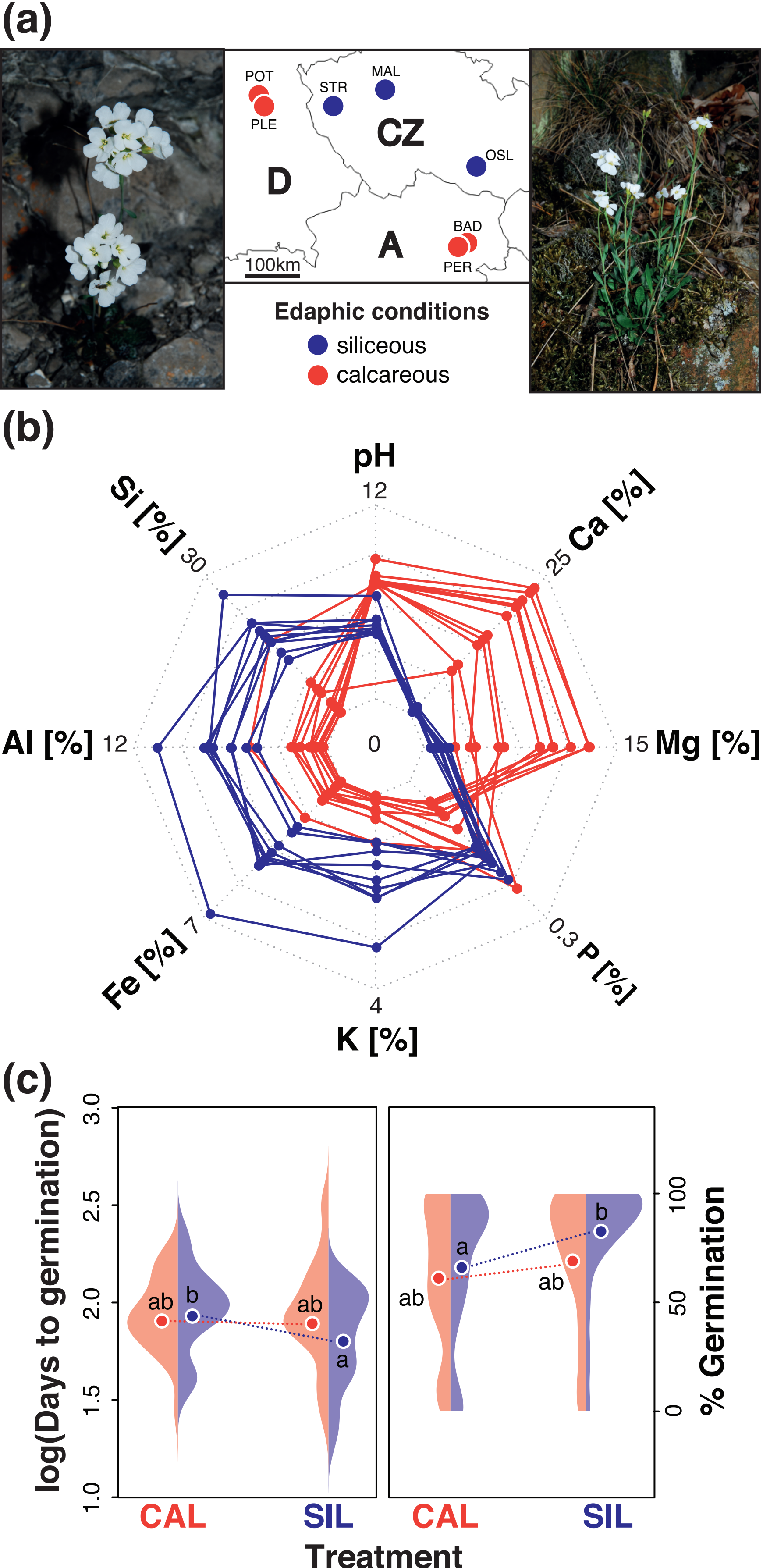
Figure 1 Sampling distribution and evidence of edaphic adaptation in *Arabidopsis lyrata*. **(a)** Geographical range of study sites for *A. lyrata* (with population acronyms) in Austria (A), Czech Republic (CZ) and Germany (D), along with photographs from a calcareous (left) and siliceous (right) population, respectively. **(b)** Radar chart displaying pH and elemental soil composition (n = 3 soil samples per study site) across three siliceous (blue) and four calcareous (red) populations of *A. lyrata*, respectively. **(c)** Mean effect of seed origin and soil treatment (CAL = calcareous, SIL = siliceous) on timing and percentage of seed germination in *A. lyrata*, along with truncated kernel density distributions estimated from data averaged by seed families for each soil treatment. Mean origin-by-treatment combinations that do not share the same letter differ significantly (n = 896 seeds, $P < 0.01$, pairwise comparisons with Bonferroni adjustment). Days to germination are shown in natural logarithm scale.

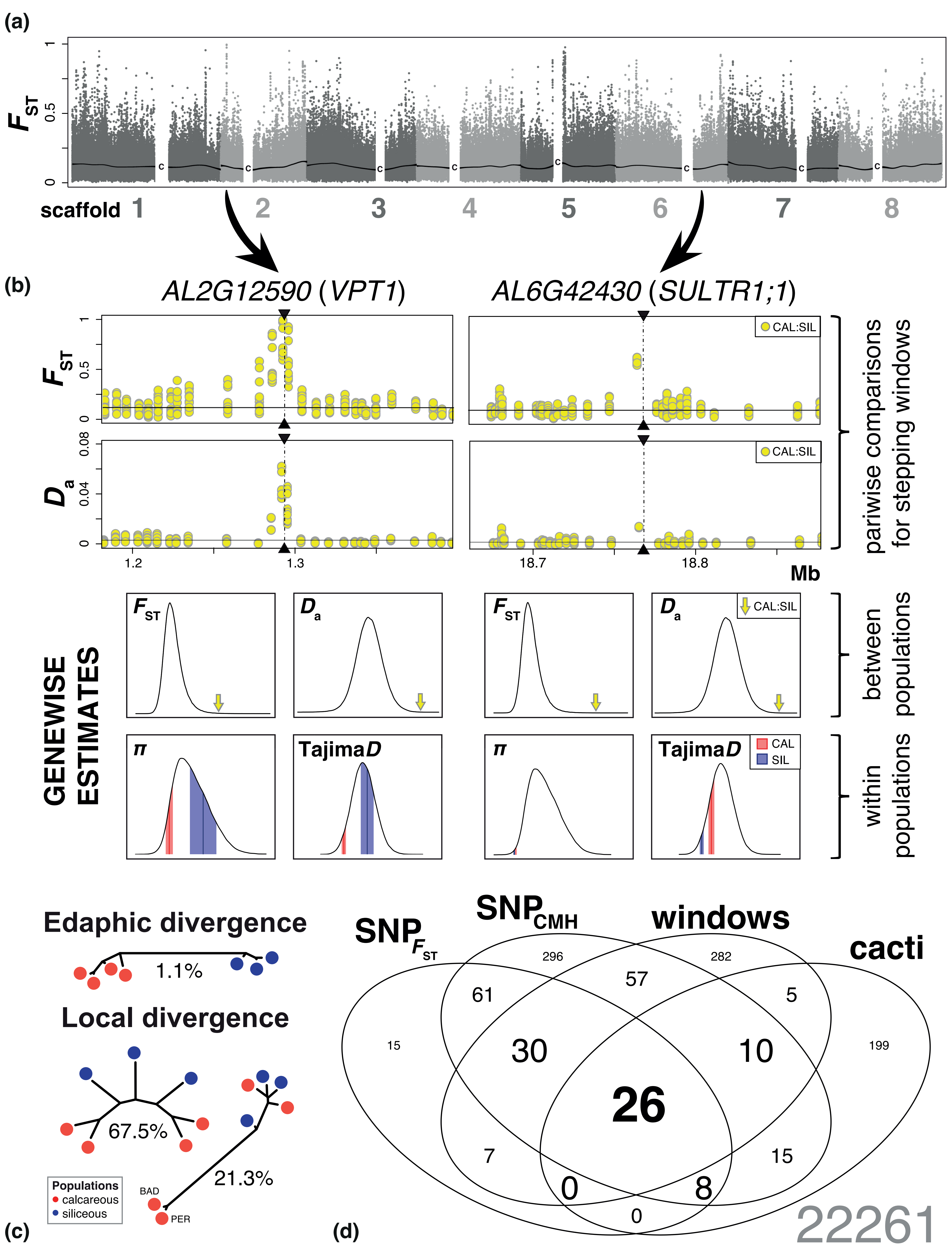
Figure 2 Evidence of divergent selection in *Arabidopsis lyrata*. **(a)** Variation of F_{ST} window estimates (light and dark grey dots corresponding to different chromosomes) for all (21) pairwise comparisons in *A. lyrata*. The 95% confidence intervals (n = 37,296 windows), calculated using a cubic smooth spline, are indicated with black lines, along with the respective location of centromeres (C). (**b**) Variation of F_{ST} and D_a window estimates for any calcareous-siliceous comparisons (upper panels) in *A. lyrata*, along with empirical distribution of standardised gene-wise F_{ST}, D_a, π and Tajima's D estimates (lower panels, n = 21,564 genes). The genomic location of homologs AL2G12590 (*VPT1*) and AL6G42430 (*SULTR1;1*) are indicated by dashed lines in the upper panels, while F_{ST} and D_a estimates are shown by yellow arrows, and mean (± SE) π and Tajima's D by shaded areas, respectively, for each given locus in the lower panels (n = 3 siliceous and 4 calcareous populations). Untransformed gene-wise estimates are provided in Supporting Information Table S7. (**c**) Patterns and proportions of local phylogenetic relationships (cacti) within the genome of *A. lyrata*. **(d)** Venn diagram summarising the number of putative adaptive gene candidates detected in *A. lyrata* using either F_{ST} SNP outlier analyses (SNP_{Fst}), CMH tests (SNP_{CMH}), sliding window approaches (windows) and/or phylogenetic inference (*viz*. cacti).

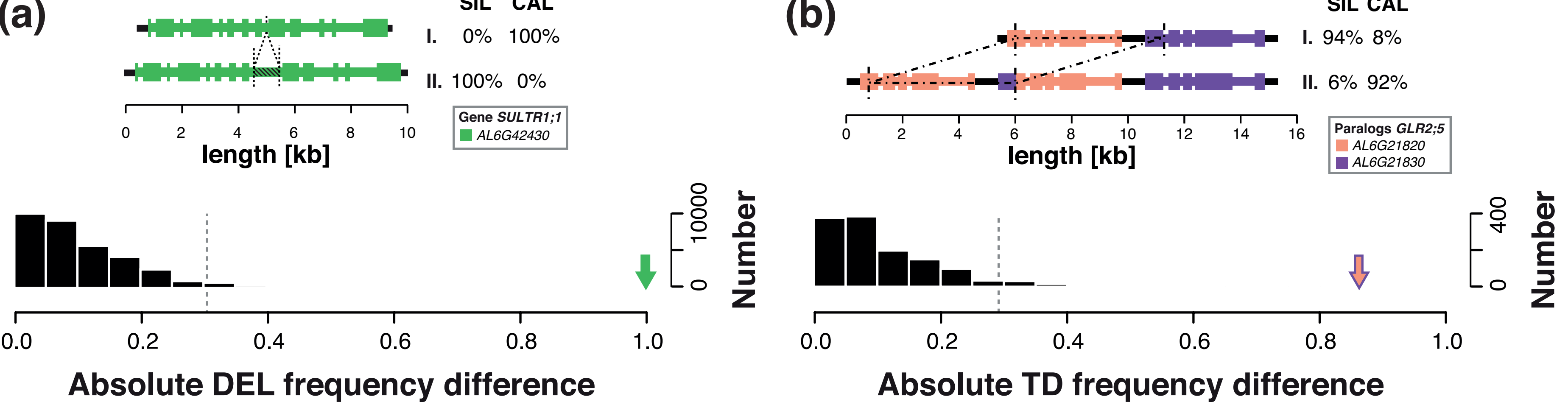
Figure 3 Examples of copy-number variants (CNVs) with high absolute frequency differences between calcareous (CAL) and siliceous (SIL) populations of *Arabidopsis lyrata*. **(a)** A deletion (DEL) in *SULTR1;1* (homolog AL6G42430) with an absolute frequency difference of 1 (green arrow), and **(b)** a tandem duplication (TD) in *GLR2;5* (paralogs AL6G21820 and AL6G21830) with highly differentiated allele frequencies (pink arrow) between edaphic ecotypes of *A. lyrata* are shown. Dotted grey lines in empirical frequency spectra delimit the 95th percentiles of absolute frequency differences for DEL (n = 40,956) and TD (n = 1,418), respectively.

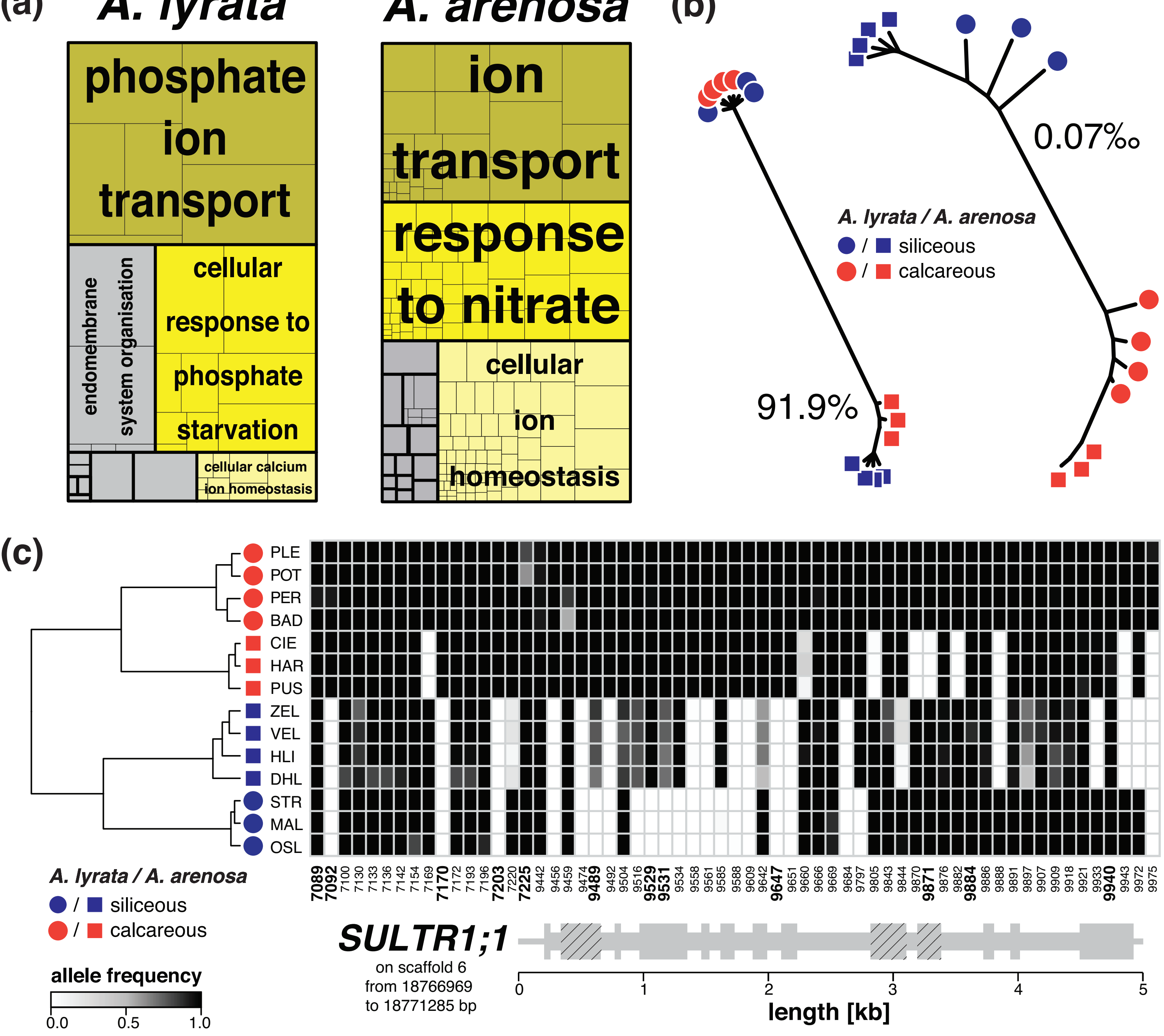
Figure 4 Evidence of shared single-nucleotide polymorphisms (SNPs) between *Arabidopsis lyrata* and *A. arenosa*. (**a**) Significantly over-represented GO terms among F_{ST} outlier SNPs detected in the respective comparisons between calcareous and siliceous populations of *A. lyrata* (n =147 genes) and *A. arenosa* (n = 2,295 genes). **(b)** Patterns and proportions of local phylogenetic relationships (*viz.* cacti) within the genome of *A. lyrata* and *A. arenosa*. **(c)** Heatmap and dendrogram of 61 SNPs shared between *A. lyrata* and *A. arenosa* at homolog AT4G08620 (*SULTR1;1*). The 12 non-synonymous polymorphisms shared across soil ecotypes are highlighted in bold.

Figure 5 Best Maximum Likelihood allelic gene tree among 72 *Arabidopsis* accessions inferred from exonic regions of *SULTR1;1* and using *Capsella rubella* as the outgroup. Bootstrap support for clades higher than 70% are indicated next to corresponding branches. Coloured bars refer to respective *Arabidopsis* lineages. The species tree inferred from whole-genome sequence data by Novikova *et al.* (2016) is shown for comparison in the bottom left corner of the figure with non-significant branching patterns collapsed along a polytomy. Acronyms next to branch tips highlight the respective calcareous (red) and siliceous (blue) populations of *A. lyrata* (circles) and *A. arenosa* (squares) investigated in this study.


(a)
POT
PLE
STR
MAL
CZ
OSL
D
A
BAD
PER
100km
Edaphic conditions
siliceous
calcareous
(b)
pH
12
Ca [%]
25
Mg [%]
15
P [%]
0.3
K [%]
4
Fe [%]
7
Al [%]
12
Si [%]
30
0
(c)
log(Days to germination)
3.0
2.5
2.0
1.5
1.0
ab
b
ab
a
CAL
SIL
a
ab
b
ab
% Germination
100
50
0
CAL
SIL
Treatment


(a)
F_{ST}
scaffold 1 2 3 4 5 6 7 8
(b)
AL2G12590 (VPT1)
AL6G42430 (SULTR1;1)
CAL:SIL
D_a
1.2
1.3
18.7
18.8
Mb
pariwise comparisons for stepping windows
GENEWISE ESTIMATES
between populations
within populations
π
TajimaD
CAL
SIL
(c)
Edaphic divergence
1.1%
Local divergence
67.5%
21.3%
BAD
PER
Populations
calcareous
siliceous
(d)
SNP$_{F_{ST}}$
SNP$_{CMH}$
windows
cacti
296
282
61
57
5
15
30
10
199
26
7
15
0
8
0
22261


(a)
SIL CAL
I. 0% 100%
II. 100% 0%
0 2 4 6 8 10
length [kb]
Gene SULTR1;1
AL6G42430
10000
0
Number
0.0 0.2 0.4 0.6 0.8 1.0
Absolute DEL frequency difference
(b)
SIL CAL
I. 94% 8%
II. 6% 92%
0 2 4 6 8 10 12 14 16
length [kb]
Paralogs GLR2;5
AL6G21820
AL6G21830
400
0
Number
0.0 0.2 0.4 0.6 0.8 1.0
Absolute TD frequency difference



(a)
A. lyrata
A. arenosa
phosphate ion transport
endomembrane system organisation
cellular response to phosphate starvation
cellular calcium ion homeostasis
ion transport
response to nitrate
cellular ion homeostasis
(b)
0.07‰
91.9%
A. lyrata / A. arenosa
siliceous
calcareous
(c)
PLE
POT
PER
BAD
CIE
HAR
PUS
ZEL
VEL
HLI
DHL
STR
MAL
OSL
A. lyrata / A. arenosa
siliceous
calcareous
allele frequency
0.0
0.5
1.0
SULTR1;1
on scaffold 6
from 18766969
to 18771285 bp
length [kb]

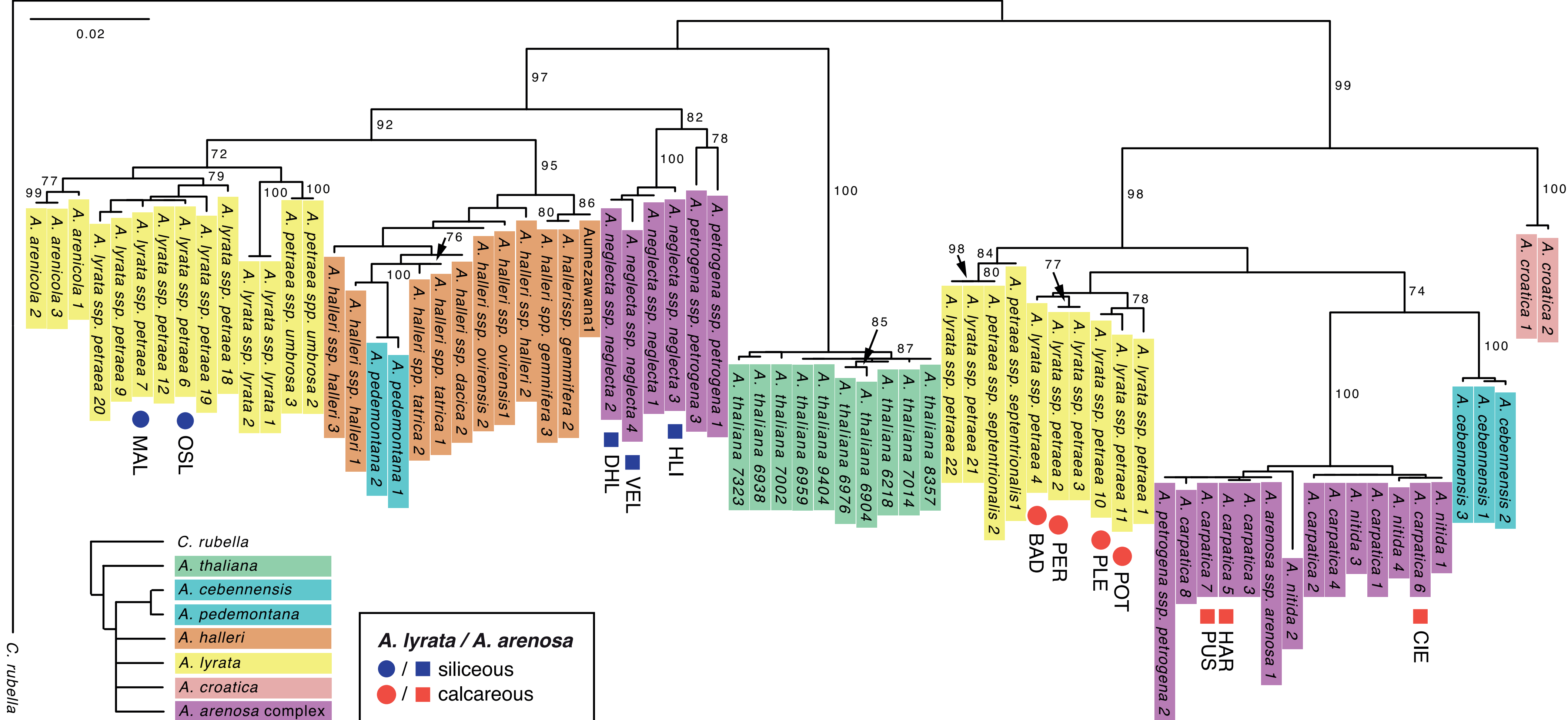

0.02
97
99
92
82
78
100
72
95
77
79
99
100
100
80
86
76
100
100
98
85
87
98
84
80
77
78
74
100
100
100
A. arenicola 2
A. arenicola 3
A. arenicola 1
A. lyrata ssp. petraea 20
A. lyrata ssp. petraea 9
A. lyrata ssp. petraea 7
A. lyrata ssp. petraea 12
A. lyrata ssp. petraea 6
A. lyrata ssp. petraea 19
A. lyrata ssp. petraea 18
A. lyrata ssp. lyrata 2
A. lyrata ssp. lyrata 1
A. petraea ssp. umbrosa 3
A. petraea spp. umbrosa 2
A. halleri ssp. halleri 3
A. halleri ssp. halleri 1
A. pedemontana 2
A. pedemontana 1
A. halleri spp. tatrica 2
A. halleri spp. tatrica 1
A. halleri ssp. dacica 2
A. halleri ssp. ovirensis 2
A. halleri ssp. ovirensis1
A. halleri ssp. halleri 2
A. halleri spp. gemmifera 3
A. hallerissp. gemmifera 2
Aumezawana1
A. neglecta ssp. neglecta 2
A. neglecta ssp. neglecta 4
A. neglecta ssp. neglecta 1
A. neglecta ssp. neglecta 3
A. petrogena ssp. petrogena 3
A. petrogena ssp. petrogena 1
A. thaliana 7323
A. thaliana 6938
A. thaliana 7002
A. thaliana 6959
A. thaliana 9404
A. thaliana 6976
A. thaliana 6904
A. thaliana 6218
A. thaliana 7014
A. thaliana 8357
A. lyrata ssp. petraea 22
A. lyrata ssp. petraea 21
A. petraea ssp. septentrionalis 2
A. petraea ssp. septentrionalis1
A. lyrata ssp. petraea 4
A. lyrata ssp. petraea 2
A. lyrata ssp. petraea 3
A. lyrata ssp. petraea 10
A. lyrata ssp. petraea 11
A. lyrata ssp. petraea 1
A. petrogena ssp. petrogena 2
A. carpatica 8
A. carpatica 7
A. carpatica 5
A. carpatica 3
A. arenosa ssp. arenosa 1
A. nitida 2
A. carpatica 2
A. carpatica 4
A. nitida 3
A. carpatica 1
A. nitida 4
A. carpatica 6
A. nitida 1
A. cebennensis 3
A. cebennensis 1
A. cebennensis 2
A. croatica 1
A. croatica 2
MAL
OSL
DHL
VEL
HLI
BAD
PER
PLE
POT
PUS
HAR
CIE
C. rubella
C. rubella
A. thaliana
A. cebennensis
A. pedemontana
A. halleri
A. lyrata
A. croatica
A. arenosa complex
A. lyrata / A. arenosa
/ siliceous
/ calcareous

Table 1 Numbers of deletions (DEL) and tandem duplications (TD) detected in *Arabidopsis lyrata*. Overall percentages of intergenic and intragenic copy-number variants (CNVs) are indicated in parentheses.

CNV Position	DEL	TD
Intergenic	26,232 (64%)	374 (26%)
Intragenic	14,724 (36%)	1,044 (74%)
Partial gene	13,125	799
Full gene	2,880	614
Total	**40,956**	**1,418**

Table 2 Genes that were consistently detected as edaphic outliers in both *Arabidopsis lyrata* and *A. arenosa*. Genes entailing shared polymorphisms are highlighted with an asterisk. na, not available.

Gene ID	TAIR	Name
AL1G17730	AT1G07550	na
AL1G18440	AT1G08080	*ACA7*
AL2G27970	AT1G68940	na
AL2G39280	AT1G78870	*UBC35*
AL5G25250*	AT3G46350	na
AL5G31760	AT3G51330	na
AL6G21820	AT5G11210	*GLR2.5*
AL6G21830	AT5G11210	*GLR2.5*
AL6G42430*	AT4G08620	*SULTR1;1*