

## ORIGINAL ARTICLE

# Local adaptation (mostly) remains local: reassessing environmental associations of climate-related candidate SNPs in *Arabidopsis halleri*

C Rellstab<sup>1</sup>, MC Fischer<sup>2</sup>, S Zoller<sup>3</sup>, R Graf<sup>1</sup>, A Tedder<sup>4</sup>, KK Shimizu<sup>4</sup>, A Widmer<sup>2</sup>, R Holderegger<sup>1,2</sup> and F Gugerli<sup>1</sup>

Numerous landscape genomic studies have identified single-nucleotide polymorphisms (SNPs) and genes potentially involved in local adaptation. Rarely, it has been explicitly evaluated whether these environmental associations also hold true beyond the populations studied. We tested whether putatively adaptive SNPs in *Arabidopsis halleri* (Brassicaceae), characterized in a previous study investigating local adaptation to a highly heterogeneous environment, show the same environmental associations in an independent, geographically enlarged set of 18 populations. We analysed new SNP data of 444 plants with the same methodology (partial Mantel tests, PMTs) as in the original study and additionally with a latent factor mixed model (LFMM) approach. Of the 74 candidate SNPs, 41% (PMTs) and 51% (LFMM) were associated with environmental factors in the independent data set. However, only 5% (PMTs) and 15% (LFMM) of the associations showed the same environment–allele relationships as in the original study. In total, we found 11 genes (31%) containing the same association in the original and independent data set. These can be considered prime candidate genes for environmental adaptation at a broader geographical scale. Our results suggest that selection pressures in highly heterogeneous alpine environments vary locally and signatures of selection are likely to be population-specific. Thus, genotype-by-environment interactions underlying adaptation are more heterogeneous and complex than is often assumed, which might represent a problem when testing for adaptation at specific loci. *Heredity* (2017) **118**, 193–201; doi:10.1038/hdy.2016.82; published online 5 October 2016

## INTRODUCTION

Local adaptation results from natural selection acting on heritable differences in fitness among individuals in response to biotic and abiotic environmental factors. As a consequence, local genotypes show higher fitness in their local habitat than genotypes originating from other habitats (Kawecki and Ebert, 2004). The genes underlying fitness-relevant heritable variation are therefore of great interest in ecology and evolution. Because next-generation sequencing (NGS) technologies allow for screening potentially millions of single-nucleotide polymorphisms (SNPs) across genomes at reasonable costs, the study of the genetic basis of local adaptation has received much attention recently (see, e.g., the reviews of Tiffin and Ross-Ibarra, 2014; Bragg *et al.*, 2015; Rellstab *et al.*, 2015). The research fields of population genomics and landscape genomics try to identify the genomic regions, genes and ultimately SNPs and their respective alleles involved in local adaptation. This is done, for example, by looking for signs of selection with outlier tests (Hohenlohe *et al.*, 2010) or environmental association analyses that correlate allele frequencies with environmental factors while accounting for confounding neutral genetic structure (Rellstab *et al.*, 2015). A widely recognized potential limitation of both outlier tests and environmental association analyses

is that many methods are known to be susceptible to false positives (e.g., de Villemereuil *et al.*, 2014; Lotterhos and Whitlock, 2015). It is therefore advised to combine approaches to reduce false positives and to assess the relevance of findings (Rellstab *et al.*, 2015).

To evaluate which candidate genes or markers are truly underlying fitness-relevant traits represents a formidable challenge. In molecular plant biology, functional validation remains the gold standard to confirm the role of a gene in controlling a specific process or trait (e.g., Weigel and Nordborg, 2005). However, functional validation is often not feasible, for example, in non-model organisms that cannot be transformed, in species having exceedingly long generation times, when the selective pressure is unknown, or when the environment cannot be simulated experimentally. Consequently, other avenues are needed to evaluate the presumed role of a candidate gene. One widely made assumption is that similar environmental conditions lead to similar genomic patterns. Hence, one can reasonably argue that identifying replicated signatures of selection in independent populations should provide strong support for an adaptive role of the identified loci (discussed, e.g., in Schmidt *et al.*, 2008; Nosil *et al.*, 2009; Prunier *et al.*, 2012; Buehler *et al.*, 2014; Tiffin and Ross-Ibarra, 2014). Such replicated patterns can derive from different processes

<sup>1</sup>Biodiversity and Conservation Biology, WSL Swiss Federal Research Institute, Birmensdorf, Switzerland; <sup>2</sup>ETH Zürich, Institute of Integrative Biology, Zürich, Switzerland; <sup>3</sup>ETH Zürich, Genetic Diversity Centre, Zürich, Switzerland and <sup>4</sup>Institute of Evolutionary Biology and Environmental Studies and Institute of Plant Biology, University of Zürich, Zürich, Switzerland

Correspondence: Dr C Rellstab, Biodiversity and Conservation Biology, WSL Swiss Federal Research Institute, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland. E-mail: christian.rellstab@wsl.ch

Received 23 March 2016; revised 5 July 2016; accepted 18 July 2016; published online 5 October 2016

(e.g., Nosil *et al.*, 2009). One possibility is that an (adaptive) allele is spread to multiple geographic localities where it is favourable and increases in frequency. Another possibility is repeated and parallel adaptation in different populations, either by independent mutations in different populations or repeated selection on standing genetic variation already present in the ancestors. A classic example for the latter is the repeated adaptation of sticklebacks (*Gasterosteus aculeatus*) to freshwater environments through the same alleles in the ectodysplasin gene, which are mostly also present in low frequencies in saltwater morphs (e.g., Jones *et al.*, 2012).

Numerous landscape genomic studies have identified candidate SNPs or genes that are presumably involved in adaptation to local environmental conditions (see, e.g., Bragg *et al.*, 2015; Rellstab *et al.*, 2015). However, most studies have not explicitly tested if the environmental associations of these candidates also hold true in independent data sets, for example, in a different region of the species' distribution range. Some studies have indirectly incorporated replicated tests by simultaneously studying adaptive genetic variation in different geographical regions and subsequently presenting the intersection of putatively adaptive loci derived from the different subsets (e.g., Poncet *et al.*, 2010; Prunier *et al.*, 2012). Interestingly, overlap among candidate loci of such population subsets is often small. For example, Poncet *et al.* (2010) found 61 and 21 climate-related amplified fragment length polymorphism loci in populations of *Arabis alpina* in the French and Swiss Alps, respectively. However, only four of these candidate loci were found in both regions studied, and only one of these showed an association to the same environmental factor. In contrast, only a few studies have tested previously identified candidates in a set of additional populations. For instance, Turner *et al.* (2010) sequenced three candidate gene regions putatively involved in adaptation to serpentine soils (identified in US populations) also in Scottish populations of *Arabidopsis lyrata*. They found the same polymorphisms involved in one and different polymorphisms within the same genomic region in the two other candidates. An important question is whether the limited overlap often observed is testimony of a poor performance (i.e., high rate of false positives) of the methods used to infer potentially adaptive variation or whether it may indicate that the genetic basis of adaptation varies among populations. Confirmation of the inferred candidate gene status may then require alternative approaches, such as transplant experiments or the analysis of patterns in nucleotide diversity.

In the present study, we tested whether putatively adaptive and climate-related SNPs in *Arabidopsis halleri* (Brassicaceae), which were identified in an earlier study (Fischer *et al.*, 2013), also show the expected environmental associations in an independent set of populations. Using whole-genome pooled sequencing (Pool-Seq) of five populations in the Swiss Alps with substantial differences in habitat characteristics, Fischer *et al.* (2013) identified over a thousand candidate SNPs in 175 genes that showed high among-population  $F_{ST}$  and that were associated with at least one of five climatic and topographic factors. In the present study, we selected a subset of 74 of these candidate SNPs located in 35 genes. First, we investigated population- and gene-specific nucleotide diversity in the original five populations to add evidence that these genes are under natural selection. Then, we tested their associations with the same environmental factors in a set of 18 additional populations comprising a larger geographic area. Our goal was to explore whether signatures of adaptation identified in a regional sample of populations reflects adaptive variation at a larger spatial scale or whether selection pressures and signatures of adaptation vary locally in heterogeneous environments such as the Swiss Alps.

## MATERIALS AND METHODS

### Study system

*Arabidopsis halleri* (L.) O'Kane and Al-Shehbaz is a perennial, insect-pollinated and strictly outcrossing herb, which is also capable of stoloniferous clonal spread. It exhibits a disjunct distribution in central Europe and eastern Asia, where it grows in various habitats including mountainous areas, grassy meadows, forest margins and rocky crevices (Clauss and Koch, 2006). A number of interesting features of this species have contributed to its development as a model plant, for example, its tolerance to and hyperaccumulation of zinc and cadmium (Clauss and Koch, 2006). The species is diploid with  $2n=16$  and is a close relative of the model species *Arabidopsis thaliana*.

### Sampling and DNA extraction

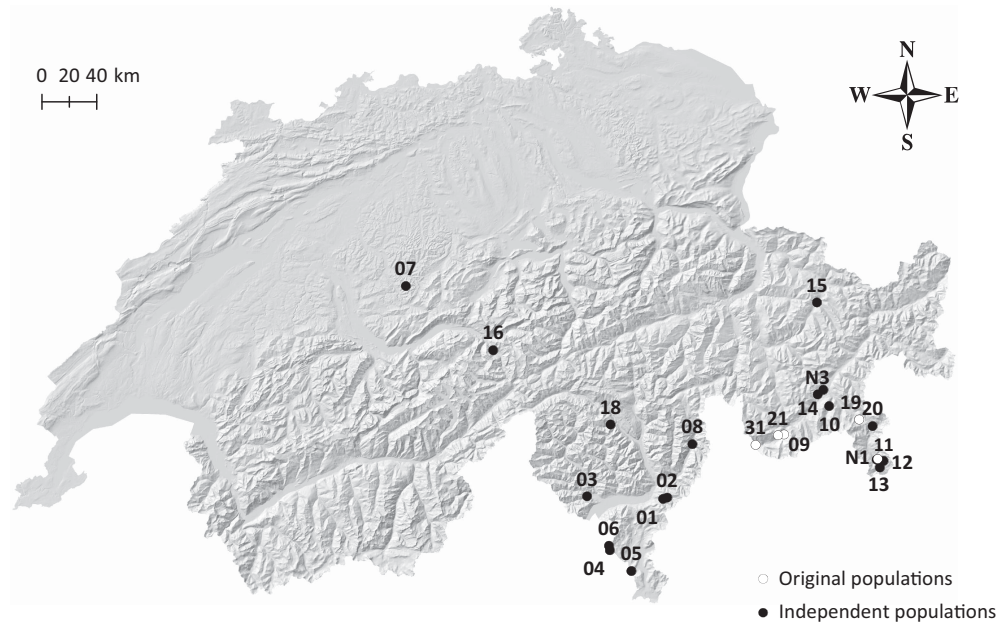
In addition to the five populations used in our previous study (Fischer *et al.*, 2013), we sampled 18 populations of *A. halleri* over a larger geographic area in pre-Alpine and Alpine regions of Switzerland (309–2305 m a.s.l.; Figure 1 and Supplementary Table S1). We collected 20 individuals per population and maintained a minimal distance of 2 m (preferably 4 m) among plants to avoid potential clones. Sampled leaf tissue was dried in silica gel, and DNA was extracted with the Qiagen DNeasy Plant Mini Kit or DNeasy 96 Plant Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

### Selection of candidate genes and SNPs

In Fischer *et al.* (2013; Table S3), 1037 SNPs in 175 genes were presented exhibiting a total of 1153 associations with one of five uncorrelated environmental factors (precipitation, slope, solar radiation, site water balance and temperature; Supplementary Table S1). In four of these genes (referred to as 'top candidate genes' in Table 2 of Fischer *et al.*, 2013), the associated environmental factors matched the biological function of the gene as based on gene ontology (GO) categories of the *Arabidopsis* Information Resource (TAIR10, <http://www.arabidopsis.org>). To select the SNPs included in this study, we applied the following procedure. The flanking sequences (50 bp on each side) of all 1037 SNPs were extracted and the major allele consensus haplotype (MACH; see Fischer *et al.*, 2013) was determined using the perl-script 'MACH.pl' (Fischer *et al.*, 2013). All resulting sequences of 101 bp length each were then quality-checked by LGC Genomics (Hoddesdon, UK) with the KRAKEN software (LGC Genomics) in view of primer design for subsequent SNP genotyping. For 730 sequences, the software was able to identify primer sites that passed the quality threshold (assay design score < 750). No assay could be designed for the only candidate SNP in the top candidate gene *ABA3* (AT1G16540). For the three remaining top candidate genes, we selected up to four SNPs per gene and associated environmental factor. For the other genes, we chose up to two SNPs per gene and associated environmental factor. The following criteria were applied for SNP selection: We excluded (i) SNPs in introns and UTR regions (except for the top candidate genes), (ii) SNPs from genes that are not functionally annotated in the *A. thaliana* reference genome (TAIR10) and (iii) SNPs that were originally associated with slope, which is difficult to interpret in view of gene functions. If possible, SNPs exhibiting multiple associations were chosen to increase the number of associations tested. Moreover, SNPs in genes with ecologically interesting functions (i.e., interpretable in relation to the environmental factors) were preferred. Finally, the primer design score of KRAKEN was also used to decide which SNPs to include in the present study. In total, we selected 79 SNPs in 35 genes (representing 91 associations with one of the environmental factors) for genotyping (Supplementary Table S2). Overall, this non-random selection of SNPs may render a successful reassessment in an independent population set more likely and simplifies the biological interpretation of our results.

### Population-specific signatures of selection

To better understand the genomic signature of selection of our candidate genes, we checked, using the original five populations and the allele frequencies of Fischer *et al.* (2013), whether we find signatures of positive selection other than (i) elevated genetic differentiation ( $F_{ST}$ ) among populations or (ii) correlation with environmental factors (as carried out in Fischer *et al.*, 2013). To do so, we analysed two population-specific estimates of gene diversity and their deviation from the exome-wide average; Tajima's  $D$  and reduction in nucleotide diversity ( $\pi$ ).



**Figure 1** Locations of the 23 populations of *Arabidopsis halleri* in Switzerland used in this study. Original populations originate from Fischer *et al.* (2013). Independent populations were additionally sampled for the present study. For detailed information see Supplementary Table S1.

We calculated exome-wide population-specific Tajima's  $D$  for 22 201 genes and  $\pi$  for 20 404 genes in the genome using the TAIR10 gene annotation and the perl-script 'Variance-at-position.pl' in POPOOLATION (Kofler *et al.*, 2011). For inferring  $\pi$ , we used a minimal coverage threshold of  $20\times$ , whereas we restricted the calculations of Tajima's  $D$  to a minimal coverage of  $13\times$ . In general, a negative genome-wide Tajima's  $D$  is indicative of an expansion after a bottleneck and negative gene-specific values reflect recent positive selection, whereas a positive genome-wide  $D$  is compatible with a decrease in population size and positive gene-specific values indicate balancing selection (Hohenlohe *et al.*, 2010). Similarly, a reduction in gene-specific  $\pi$  compared with the average genome-wide level shows that positive selection most likely has acted on this gene within this population. A reduction in  $\pi$  indicates more recent selection, whereas negative Tajima's  $D$  values refer to a slightly older signature of selection (Hohenlohe *et al.*, 2010). To detect rather strong signals of selection, we considered a gene in a specific population under positive selection when Tajima's  $D$  was below  $-1$  or the reduction in  $\pi$  was at least twofold as compared with the average exome-wide level in the given population.

### SNP genotyping

In Fischer *et al.* (2013), SNP allele frequencies were derived from sequencing (Illumina HiSeq2000) of population pools (each containing equimolar quantities of DNA from 20 individuals) and subsequent mapping of the resulting paired-end reads to the *A. thaliana* reference genome. For the present study, genotyping of the 79 selected SNPs was performed by LGC Genomics using the Kompetitive Allele-Specific PCR (KASP) technology for all 23 populations, that is, including the five original populations. The KASP genotyping assays contain three assay-specific unlabelled primers (Supplementary Table S2); two allele-specific forward primers and one common reverse primer.

### Neutral genetic population structure

In Fischer *et al.* (2013), neutral genetic population structure was characterized by calculating the average pairwise  $F_{ST}$  using more than two million SNPs across the *A. halleri* genome. For this study, we applied eight nuclear simple sequence repeat (nSSR, microsatellite) markers specifically developed for *A. halleri* by Godé *et al.* (2012) to estimate pairwise  $F_{ST}$  values among all 23 populations. Multiplex PCR reactions were conducted with the Qiagen Type-it Microsatellite PCR Kit using labelled forward primers (Microsynth, Balgach, Switzerland). We used a total reaction volume of  $10\ \mu\text{l}$  containing  $4\ \mu\text{l}$

of Multiplex PCR master mix,  $2.7\ \mu\text{l}$  of primer mix ( $2\ \mu\text{M}$  of each primer) and on average  $3\ \text{ng}$  DNA in the remaining  $3.3\ \mu\text{l}$  RNase-free water. The cycling protocol consisted of 5 min initial denaturation at  $95\ ^\circ\text{C}$ , 31 cycles of 30 s denaturation at  $95\ ^\circ\text{C}$ , 90 s primer annealing at  $58\ ^\circ\text{C}$  and 30 s extension at  $72\ ^\circ\text{C}$  and a final extension of 45 min at  $65\ ^\circ\text{C}$ . We then mixed  $1\ \mu\text{l}$  of  $10\times$  diluted PCR product with  $10\ \mu\text{l}$  of a formamide/size standard mixture ( $10\ \mu\text{l}$  GeneScan-500 LIZ size standard in  $1\ \text{ml}$  HiDi formamide, both from Applied Biosystems, Waltham, USA) for fragment analysis on an ABI 3730 DNA Analyser (Applied Biosystems). Allele scoring was semi-automatically performed using GENEMAPPER 5 (Applied Biosystems). Null allele frequencies and pairwise  $F_{ST}$  among populations were calculated with GENEPOP 4.2.2 (Rousset, 2008).

### Environmental association analysis

Hereafter, we refer to three different data sets as follows (see also Figure 1): the five populations genotyped on the basis of Pool-Seq data (Fischer *et al.*, 2013) represent the 'original data set'; the same five populations genotyped individually with KASP and nSSRs (this study) form the 'reference data set'; the 'independent data set' consists of the 18 newly sampled populations genotyped individually with KASP and nSSRs (this study). We aimed to test whether the candidate SNP loci of the independent data set can be associated with the same environmental factors as in the original data set. The reference data set was only used for assessing the effect of the different genotyping methods (Pool-Seq vs individual genotyping with KASP/nSSRs) by comparing it to the original data set. Note that the original data set only includes results from Fischer *et al.* (2013) without any additional analyses.

We used the same statistical approach (partial Mantel tests, PMTs) and environmental factors (precipitation, slope, solar radiation, site water balance and temperature) as in Fischer *et al.* (2013) to associate loci with the environment. The fact that we did not select any SNPs originally associated with slope and that we still used slope in all environmental association analyses does not affect our main conclusions, because we were mainly interested in the proportion of confirmed associations in the independent data set. In contrast to the original study, we inferred individual genotype information instead of relying on allele frequency estimates from Pool-Seq. As PMTs have repeatedly been criticized in the past (e.g., Guillot and Rousset, 2013), we also analysed the independent data set with latent factor mixed models (LFMMs; Fricot *et al.*, 2013) that are furthermore better suited to deal with individual genotype data.



Data of the environmental factors used (Supplementary Table S1) were extracted for each sampling location using ARCMAP 10 (ESRI) from interpolated GIS data at 25-m resolution collected by a dense network of meteorological stations over a 30-year period (1961–1990; Zimmermann and Kienast, 1999). Except for environmental data measured on-site, which is rarely available for a given set of sampled populations, this represents the best-resolved data available for environmental association studies in natural populations.

PMTs examine correlations between two pairwise distance matrices, while controlling for the effect of a third, and have been widely used in environmental association analyses (Rellstab *et al.*, 2015). Here, the dependent variable was the matrix of pairwise population  $F_{ST}$  values at a given SNP, the predictor variable was the corresponding matrix of pairwise distances of a given environmental factor, and the third matrix contained the pairwise population  $F_{ST}$  values to account for neutral genetic structure. We used GENEPOP 4.2.2 to calculate pairwise  $F_{ST}$  for each SNP locus. We did not use the genotyped SNPs to calculate neutral  $F_{ST}$ , as the candidate SNPs were expected to be non-neutral. Instead, we incorporated the pairwise  $F_{ST}$  values from the eight nSSR markers described above. The use of  $F_{ST}$  values derived from nSSRs to correct for neutral genetic structure of SNPs is debated (Rellstab *et al.*, 2015). However, a previous analysis in *A. halleri* showed that there is a strong correlation between genome-wide  $F_{ST}$  using SNPs and  $F_{ST}$  with the same nSSR markers applied here (Mantel  $r=0.866$ ; MC Fischer, unpublished results). As PMTs are not dependent on the scale of the matrices, the fact that  $F_{ST}$  values using nSSR markers were consistently higher compared with genome-wide  $F_{ST}$  values (3.4x; MC Fischer, unpublished results) can be ignored. PMTs were performed with ECODIST 1.2.7 (Goslee and Urban, 2007) in R 3.2.1 (<http://www.r-project.org>) with 1001 permutations. As in Fischer *et al.* (2013), we used the effect size ( $r$ ) as a threshold criterion. To estimate these thresholds for the reference and the independent data sets, we performed PMTs between 16 383 matrices with simulated pairwise  $F_{ST}$  values (random values between 0 and 1) and the five environmental factors (real values), again controlling for neutral population structure (real pairwise  $F_{ST}$  from nSSRs). We used the 99% quantile of the resulting 81 915  $r$  values as threshold. For the reference data set, this threshold was 0.7557, for the independent data set, it was 0.1890. Above these thresholds, a specific SNP locus was considered to be significantly associated with the respective environmental factor. The large difference in the threshold values derives from the different numbers of pairwise comparisons (10 in the reference and 153 in the independent data set).

The software LFMM (Frichot *et al.*, 2013) uses a linear mixed model to test for associations between genetic variation (response variable) and environmental factor (explaining variable), while controlling for neutral genetic structure with (random) latent factors. It has been shown to be an efficient method for different scenarios of demographic history (e.g. de Villemereuil *et al.*, 2014; Lotterhos and Whitlock, 2015). We ran LFMM 1.4 with the independent data set using 1000 sweeps for burn-in and 9000 additional sweeps, testing environmental factors one by one. We performed 10 runs per environmental factor, but as run-to-run variation was extremely low (minimum Pearson's  $r=0.999$  for all pairwise comparisons of  $Z$  values), we only kept the first run for further analysis. We then chose significant associations based on a false rate of 5% ( $q \leq 0.05$ ) using the R package QVALUE (Storey and Tibshirani, 2003). As suggested by the authors of LFMM (Frichot *et al.*, 2013), we used the number of population clusters as the numbers of latent factors ( $K_{LFMM}$ ). To determine the number of clusters, we ran STRUCTURE 2.3.3 (Pritchard *et al.*, 2000) with the eight nSSR markers, using 10 simulations with different seeds for each  $K_{STRUCTURE}$  from 1 to 18 using 1 000 000 repetitions after a burn-in period of 100 000 runs, admixture model, correlated allele frequencies and no prior population information. We chose the  $K_{STRUCTURE}$  at which the mean probability of the data reached a plateau, as recommended by Pritchard *et al.* (2000). We verified our choice of  $K_{LFMM}$  based on histograms of test  $P$  values (François *et al.*, 2016).

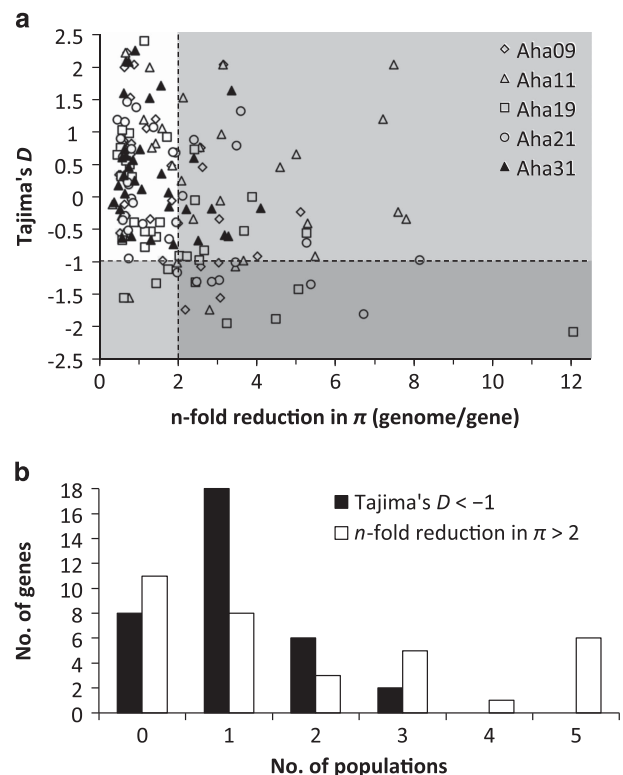
To assess the direction of the association (whether allele frequencies decrease or increase with an increasing environmental factor) and for illustration, we also performed linear regressions between allele frequencies at candidate SNPs and the five environmental factors in R for all three data sets. If a significant association was found in the original and independent data set, we checked whether the sign of the regression slope was identical in the two data sets. Only

associations with the same sign of the regression slope in both data sets were considered confirmed.

## RESULTS

### Population-specific signatures of selection

Our analyses of Tajima's  $D$  and reduction in  $\pi$  revealed that 27 (79%) of 34 genes (no calculation was possible for *MIR398C* (AT5G14565)) showed population-specific signatures of selection in at least one of the five original populations for at least one of the two parameters considered (Figure 2a). In 22 genes (65%), this was the case for both parameters. Average exome-wide Tajima's  $D$  was between  $-0.021$  (population Aha21) and  $-0.119$  (Aha31). In 26 genes, at least one population showed a Tajima's  $D$  below  $-1$ , which we defined as strong evidence for selection (Figure 2b). In 18 of these genes, evidence for selection was found only in a single population. Average exome-wide  $\pi$  varied between 0.008 (Aha11) and 0.009 (Aha31). Reduction in  $\pi$  (in comparison with exome-wide levels) was on average 2.2-fold across the 34 genes. 23 of the 34 genes showed an at least twofold reduction in  $\pi$  in at least one population (Figure 2b), which we considered as a strong signal of selection. In eight of these cases, only a single population showed a signature of selection.



**Figure 2** Population- and gene-specific signatures of selection estimated for the five populations of the original data set, calculated from exome-wide Pool-Seq data from Fischer *et al.* (2013). (a) Reduction in gene-specific nucleotide diversity ( $\pi$ ) compared with the exome-wide  $\pi$  and gene-specific Tajima's  $D$  for each of the 34 analysed genes in each population. Values above 2 in reduction in  $\pi$  and below  $-1$  in Tajima's  $D$  are considered of a gene as being under strong positive selection in the specific population. (b) Number of genes in the number of populations in which these genes show signs of selection given the criteria above.

### SNP genotyping

Of the 79 SNP assays, four failed to amplify despite high assay design quality scores, and one assay showed no variation, most likely because of null alleles (Supplementary Table S2). Four samples were excluded, because they could only be scored at 10 or less of the 74 SNP loci. In total, we used 444 samples for further analyses (reference data set 98 samples, independent data set 346 samples); genotyping success rate was 96.9% in these individuals. First, as in Rellstab *et al.* (2013), we checked whether the pooled approach (original data set, using Illumina whole-genome sequencing and mapping to the *A. thaliana* reference genome) of Fischer *et al.* (2013) could be validated using individual genotyping with KASP (reference data set), that is, whether both methods yielded the same allele frequencies per population and SNP. We found that the two methods showed a very good match (linear regression,  $n=370$ ,  $R^2=0.954$ ,  $P<0.001$ ), with a slope of 1 and an intercept of 0; Supplementary Figure S1a). Average difference in allele frequency per SNP and population between the two approaches was 4.5%. However, in some cases, the difference in allele frequency was substantially higher; in 37 and 6 out of 370 comparisons, the difference was above 10% and 20%, respectively.

### Neutral genetic structure

All samples could be successfully genotyped in at least six nSSR loci. In 86.3% of samples, all eight nSSR markers amplified. No marker exhibited substantially increased null allele frequencies across all populations; only in 36 out of 192 cases, the estimated null allele frequency was above 10%. Pairwise  $F_{ST}$  values (Supplementary Table S3) among populations ranged between 0.003 (Aha10 vs Aha19) and 0.587 (Aha01 vs Aha15).  $F_{ST}$  values inferred from Pool-Seq (original data set) and nSSRs (reference data set) showed moderate congruence (Mantel test with Pearson's  $r$ ,  $n=10$ ,  $r=0.601$ ,  $P=0.168$ ; Supplementary Figure S1b).

### Environmental association analyses

Using PMTs, we found 53 associations for 30 SNPs located in 16 genes (Table 1) in the independent data set (highest partial Mantel  $r=0.605$ ). Temperature (16 associations) and precipitation (14) were most frequently associated with candidate SNPs, followed by site water balance (8), radiation (8), and slope (7); for details see Supplementary Table S2. In the reference data set (highest partial Mantel  $r=0.971$ ), we found 102 associations for 73 SNPs with an environmental factor. Of these, 70 were also present in the list of 85 associations of the

original data set. We found a good match (linear regression,  $n=370$ ,  $R^2=0.774$ ,  $P<0.001$ ) between the  $r$  values of the partial Mantel tests of the original and reference data sets (i.e., Pool-Seq. vs KASP/nSSRs; Supplementary Figure S1c). However, in some comparisons,  $r$  values differed substantially as a cumulative effect of differences in allele frequencies (Supplementary Figure S1a) and neutral  $F_{ST}$  values (Supplementary Figure S1b) generated by the two different methodological approaches.

Before the LFMM analyses, the number of latent factors ( $K_{LFMM}$ ) had to be chosen. In STRUCTURE, the mean probability of the data plateaued at  $K_{STRUCTURE}=9$  (Supplementary Figure S2). The non-significant  $P$  values with  $K_{LFMM}=9$  were more or less uniformly distributed (Supplementary Figure S3), ensuring an efficient control of the false discovery rate (François *et al.*, 2016). Therefore,  $K_{LFMM}=9$  was chosen as the number of latent factors. The LFMMs using a false discovery rate of 5% (corresponding to a threshold of  $|Z|=2.385$ ) found 76 significant associations for 38 SNPs located in 23 genes in the independent data set (Table 1). Again, temperature (20) and precipitation (18) exhibited the highest numbers of associations, followed by slope (13), site water balance (13) and radiation (12); for details see Supplementary Table S2. The  $|Z|$  values of LFMMs and the  $r$  values of PMTs were weakly congruent (linear regression,  $n=370$ ,  $R^2=0.176$ ,  $P<0.001$ ; Supplementary Figure S4); 25 of the 53 associations (47.2%) identified using PMTs were also confirmed by LFMMs. When considering only the associations identified by both approaches,  $|Z|$  and  $r$  values were more congruent (linear regression,  $n=25$ ,  $R^2=0.289$ ,  $P<0.001$ ).

As 11 of the 74 tested SNPs were found to be associated with two environmental factors in the original data set (Fischer *et al.*, 2013), we aimed at testing 85 candidate associations; 19 associations with precipitation, 24 with radiation, 27 with site water balance and 15 with temperature. Of these, we found 7 (PMTs) and 17 (LFMMs) associations also in the independent data set. However, in three (PMTs) and six (LFMMs) cases, the signs of the slopes of the independent data set were contrary to those of the original data set (see, e.g., Supplementary Figure S5), resulting in 4 (4.7%, PMTs) and 11 (15.3%, LFMMs) environmental associations identically found in the original and the independent data sets (Table 1). Two of these associations were detected with both association methods (Figure 3 and Supplementary Figure S6). Table 2 presents the 13 candidate SNPs located in 11 genes that showed the same significant associations in the 5 original and 18 independent populations using PMTs and/or LFMMs.

**Table 1** Number of genes, SNPs and associations identified in *Arabidopsis halleri* using different environmental association approaches and data sets

Environmental association analysis and data set	No. of genes <sup>a</sup>	No. of SNPs <sup>a</sup>	No. of associations <sup>a</sup>
Candidates identified (OD <sup>b</sup> )	175	1037	1153
Candidates tested (ID)	35	74	85
PMTs (ID)	16 (44%)	30 (41%)	53
PMTs (OD∩ID)	4 (11%)	4 (5%)	4 (5%)
LFMMs (ID)	23 (66%)	38 (51%)	76
LFMMs (OD∩ID)	9 (26%)	11 (15%)	11 (13%)
PMTs and LFMMs (OD∩ID)	11 (31%)	13 (18%)	13 (15%)

Abbreviations: ID, independent data set; OD, original data set; PMTs, partial Mantel tests; LFMMs, latent factor mixed models; ∩, overlap between data sets; SNP, single-nucleotide polymorphism.

<sup>a</sup>The proportions given refer to the candidates tested in the ID.

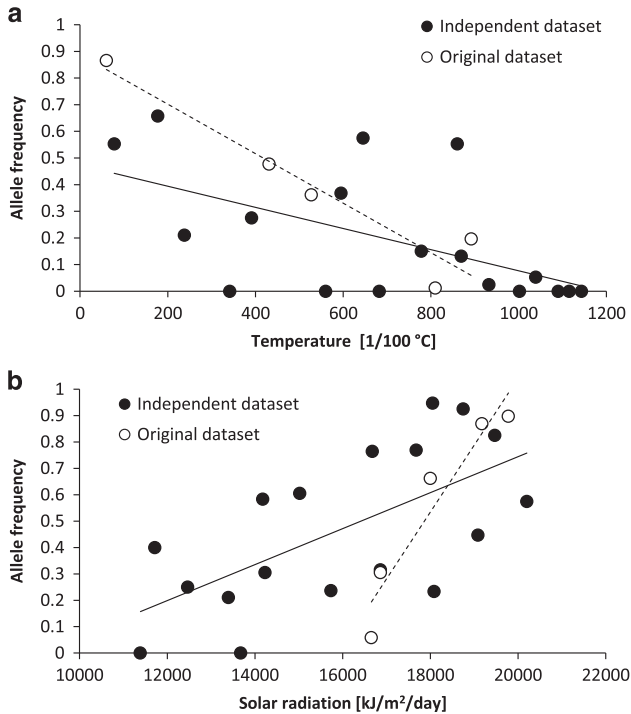
<sup>b</sup>Candidates that were reported in the OD (Fischer *et al.*, 2013).

### DISCUSSION

Among others, landscape genomic studies aim at identifying genes that underlie adaptation to local environmental conditions. However, due to various methodological and biological reasons (for a discussion, see Rellstab *et al.*, 2015), such analyses can result in fairly large numbers of false-positive findings. It is therefore recommended to increase the number of true positives by adding additional evidence from alternative approaches. One possibility is to search for replicated adaptive patterns in independent data sets, as repeatedly finding the same candidate loci is strong evidence that these loci are actually involved in adaptation to certain environmental factors (Schmidt *et al.*, 2008; Prunier *et al.*, 2012; Tiffin and Ross-Ibarra, 2014). However, not all candidate loci that cannot be confirmed in independent population data sets represent false positives. An alternative explanation is that selection pressures differ in different regions of the species' natural range, leading to different patterns of adaptive genetic variation (e.g., Feulner *et al.*, 2015). Thus, comparative analyses in different

sets of populations may reveal interesting insight into the relative amounts of locally restricted and parallel adaptive processes.

We approached this topic by reassessing environmental associations in a subset of putatively adaptive and climate-related loci of the



**Figure 3** Allele frequencies vs environmental factors for the two SNPs that were successfully confirmed in both partial Mantel tests and latent factor mixed models in *Arabidopsis halleri*. (a) Locus Chr1\_2139123 from gene *CHX14* (TAIR ID AT1G06970) associated with temperature. (b) Locus Chr1\_19775226 from a *protein kinase superfamily protein* (AT1G53050) associated with solar radiation. Fitted lines show linear regressions (dashed line: original data set; solid line: independent data set). Linear regressions are here only for illustration (they do not account for neutral genetic structure and were not used in the environmental association analysis; the residual plots of the partial Mantel tests are shown in Supplementary Figure S6).

Brassicaceae *Arabidopsis halleri*. To do so, we genotyped SNPs, derived from candidate genes identified by Fischer *et al.* (2013), in a set of 18 additional populations from a larger geographic area as compared with the original study. Irrespective of the association approach used, the environmental association of a majority of the 74 candidate SNPs could not be confirmed in the independent data set (Table 1). Although the independent data set resulted in 53 associations (41% of the SNPs) of a SNP with an environmental factor using PMTs, only four (5%) of these associations overlapped with the original data set. The success rate marginally improved when LFMMs were used; 38 SNPs (51%) showed an association in the independent data set, and 11 (13%) of these associations matched with those detected in the original data set.

Cumulatively (PMTs and LFMMs), there were 13 SNPs, corresponding to 11 genes, that showed consistent associations in the original and independent data set (Table 2). These genes can thus be considered candidate genes for environmental adaptation at a broader geographical scale. Four of the 13 confirmed SNPs, located in four genes, were found using the same method as in the original data set (PMTs). One SNP located in the gene *CHX14* (AT1G06970), a plasma membrane potassium (K<sup>+</sup>) efflux transporter involved in K<sup>+</sup> homeostasis and recirculation from root to leaf tissues (Zhao *et al.*, 2015), was associated with temperature. In goatgrass species (*Aegilops* spp.), K<sup>+</sup> concentration and leakage were associated with drought tolerance (Farooq and Farooq, 2001), suggesting a link between the maintenance of K<sup>+</sup> transport and temperature. Another SNP was associated with solar radiation and is located in a poorly described gene (AT1G53050) belonging to the protein kinase superfamily involved in signal-transduction pathways. The third SNP, associated with site water balance, is located in the gene *SAR3* (AT1G80680). The gene family to which *SAR3* belongs is known to have a role in the suppression of the transcriptional repressors of the *AXR* gene family (Parry *et al.*, 2006), which was found to have a regulatory role for various proteins in *Arabidopsis* lines undergoing progressive drought stress (Bianchi *et al.*, 2002), indicating a putative link between this gene and the associated climatic factor. Finally, one SNP in the gene *GRL3.6* (AT3G51480) was also associated with site water balance. This gene is potentially involved in light-signal-transduction (Lacombe *et al.*, 2001) and leaf-to-leaf wound signalling (Mousavi *et al.*, 2013). However, contrary to the original data set (Fischer *et al.*, 2013), the additional association of

**Table 2** Environmental associations successfully confirmed by either PMTs and/or LFMMs in the independent population data set of *Arabidopsis halleri*

SNP	<i>r</i> (PMTs) <sup>a</sup>	IZI (LFMMs) <sup>a</sup>	Associated with	Chromosome	Position	TAIR ID	Primary gene name
2	-0.080	<b>2.738</b>	Precipitation	Chr1	1 710 422	AT1G05700	<i>Leucine-rich repeat transmembrane protein kinase protein</i>
4	<b>0.266</b>	<b>3.141</b>	Temperature	Chr1	2 139 123	AT1G06970	<i>CHX14—cation/hydrogen exchanger 14</i>
10	<b>0.254</b>	<b>2.587</b>	Solar radiation	Chr1	19 775 226	AT1G53050	<i>Protein kinase superfamily protein</i>
17	<b>0.500</b>	0.632	Site water balance	Chr1	30 325 655	AT1G80680	<i>SAR3—suppressor of auxin resistance 3</i>
20	-0.116	<b>3.588</b>	Solar radiation	Chr2	13 940 829	AT2G32860	<i>BGLU33—β-glucosidase 33</i>
44	-0.033	<b>2.871</b>	Temperature	Chr3	6 573 965	AT3G19040	<i>HAF2—histone acetyltransferase of the TAFII250 family 2</i>
51	0.163	<b>2.451</b>	Solar radiation	Chr3	9 547 624	AT3G26120	<i>TEL1—terminal ear1-like 1</i>
56 <sup>b</sup>	<b>0.346</b>	1.936	Site water balance	Chr3	19 101 737	AT3G51480	<i>GRL3.6—glutamate receptor 3.6</i>
64	0.153	<b>4.086</b>	Site water balance	Chr4	11 250 166	AT4G21070	<i>BRCA1—breast cancer susceptibility 1</i>
69	0.166	<b>2.995</b>	Precipitation	Chr5	22 937 358	AT5G56680	<i>SYNC1</i>
70	0.065	<b>3.040</b>	Precipitation	Chr5	22 937 500	AT5G56680	<i>SYNC1</i>
73	0.144	<b>2.413</b>	Temperature	Chr5	25 328 983	AT5G63140	<i>PAP29—purple acid phosphatase 29</i>
74	0.162	<b>3.754</b>	Temperature	Chr5	25 329 190	AT5G63140	<i>PAP29—purple acid phosphatase 29</i>

Abbreviations: LFMMs, latent factor mixed models; PMTs, partial Mantel tests; SNP, single-nucleotide polymorphism.

<sup>a</sup>Significant values are shown in bold. Significance threshold for PMTs: *r* > 0.189. Significance threshold for LFMMs: IZI > 2.385.

<sup>b</sup>Top candidate gene from Fischer *et al.* (2013).



this gene with solar radiation was just below the defined association threshold (partial Mantel  $r=0.180$ ; LFMM  $|Z|=1.936$ ). Moreover, LFMMs identified nine SNPs located in seven genes that were associated with the same environmental factor as in the original data set (Table 2). Of these, the gene *SYN1* (AT5G56680) showed an interesting link between its function and the associated factor; it is involved in seed dormancy and contained two SNPs associated with precipitation.

Methodological differences between the original (Fischer *et al.*, 2013) and the present study might be partly responsible for the fact that we could only identify four (5%) of the originally 85 associations using PMTs. The main difference is the genotyping approaches. Whereas allele frequencies showed largely good congruence between the two approaches (Pool-Seq vs KASP; Supplementary Figure S1a), the characterization of neutral genetic structure differed (whole-genome SNPs vs nSSRs; Supplementary Figure S1b). However, our comparisons of the partial Mantel  $r$  values between original and reference data sets (Supplementary Figure S1c) and the fact that 70 of the 85 tested associations could also be found in the reference data set indicate that the differences between the two methods could not sufficiently explain the low number of associations retrieved in the independent data set. Another methodological difference is that the 74 loci tested in the present study are not necessarily highly differentiated SNPs in terms of  $F_{ST}$ , which was a precondition in the original study. However, average maximum allele frequency differences among populations indicated that many loci of the independent data set were similarly differentiated ( $0.786 \pm 0.200$  s.d.) as those of the original data set ( $0.807 \pm 0.060$ ). Only 20 of 74 SNPs in the independent data set exhibited a maximum allele frequency difference below the minimum (0.707) of the original data set.

It is a general limitation of environmental association analyses that many (also unmeasured) environmental factors are often correlated with each other (Rellstab *et al.*, 2015). Consequently, relevant associations could be missed because a highly correlated factor functioned as the causal driver (or a proxy for it) behind patterns of allele frequency, but was not included in the analysis. In our specific case, we used five environmental factors derived from a set of 21 factors, after removing highly correlating factors (Fischer *et al.*, 2013). Pairwise correlations among the five environmental factors differed substantially between the original and independent data sets (Supplementary Table S4). We therefore checked if additional associations could be found in cases where the associated factors between the independent and the original data sets did not match, but were mediated through a third correlated factor. In the case of PMTs, we found five such cases, all concerning moisture index in summer (the difference between precipitation and potential evaporation from June to August), which was highly correlated with precipitation and site water balance in the original data set and additionally with temperature in the independent data set. However, in only one of the five loci, the regression of moisture index and allele frequency pointed into the same direction in the original and independent data set. This was an SNP (number 63 in Supplementary Table S2) in the gene *ATSPSAF* (AT4G10120), which was correlated with site water balance in the original data set and temperature in the independent data set. This gene has a regulatory function in sucrose synthesis (Volkert *et al.*, 2014). The number of missed associations through originally excluded correlated factors therefore seemed to be negligible in our case.

The low congruence between the results of the original and the independent data sets might also derive from the characteristics of the original data set (Fischer *et al.*, 2013). The study was based on a relatively low number of populations, which might have led to limited

generalization regarding candidate SNPs and environmental associations. The associations from the original data set might have derived from environments and genetic backgrounds that only covered a part of the range that *A. halleri* actually experiences. For example, solar radiation (Figure 3b) in the original data set was in the upper range of that of the independent data set. The possibility that some of the candidate genes of Fischer *et al.* (2013) represent false positives cannot be completely excluded. However, our analysis of population-specific signatures of selection (Figure 2) revealed that the tested candidate genes mostly showed strong signals of selection in at least one population of the original data set.

Because we can exclude substantial effects of methodological differences on the outcome of our study, our results suggest that most environmental associations identified in either the original or independent data sets are truly local. The genotype-by-environment interactions that are modulating selection seem to be much more complex than is often assumed and may create genomic selection patterns in an unpredictable manner, leading to geographically restricted local adaptation (Schmidt *et al.*, 2008). For example, many adaptive processes are based on polygenic selection (Pritchard and Di Rienzo, 2010) and might result in different patterns of allele frequencies depending on the environmental and genetic backgrounds of respective populations. Moreover, different molecular mechanisms might be responsible for convergent phenotypic evolution of populations (as for example the Zn tolerance in *A. halleri*; Pauwels *et al.*, 2006). In the complex landscape of genetic and environmental variation, it might thus be difficult to find environmental associations that hold true in general, especially in heterogeneous environments such as those encountered in topographically rugged alpine habitat. Similar evidence has been presented in an overview of landscape genomic studies of long-lived forest trees exhibiting high gene flow (Čalić *et al.*, 2015). This shows that our findings may also hold true in different biological systems. This locally restricted effect of natural selection and adaptation is also partly confirmed by the population-specific signatures of selection in the original data set. The effects of selection were often apparent in only one or few populations (Figure 2b).

Our results have implications for studies investigating the genetic basis of adaptation to the environment. They show that the geographical extrapolation of a candidate locus has to be explicitly tested. However, such studies have been, to our knowledge, very scarce and/or unsuccessful. For example, Buehler *et al.* (2014), using 30 independent populations of *Arabis alpina*, failed to validate an outlier locus associated with different habitat types. In beach mouse, allele frequencies underlying the same patterns of coat colour differed in separate geographic regions (Hoekstra *et al.*, 2006), indicating alternative loci or genes affecting fur colour. In 136 accessions of *A. thaliana* from across Europe, Korves *et al.* (2007) could not find the well-known association of certain genotypes of the *FRIGIDA* gene with seasonal environment and flowering time. The search for parallel adaptive patterns in different geographical regions is common in speciation genomics. In sticklebacks, it has repeatedly been shown that the same loci are responsible for the parallel evolution of distinct marine and freshwater sticklebacks (e.g., Jones *et al.*, 2012). In contrast, lake and river morphs do not seem to show much parallelism. Feulner *et al.* (2015), for example, showed that outlier gene regions among pairs of lake vs river populations were locally specific. Although these studies do not represent classical landscape genomic approaches, they give interesting insights into the level and magnitude of parallel and local adaptation.

When should one expect analyses to provide general insights when they are based on a limited number of populations? First of all, the populations investigated should ideally cover the range of environmental conditions that the study species occupies. If this is not the case, the combination of low sample size and limited environmental range may lead to false positives, because the associations between allele frequencies and environment can take different directions and forms (as shown for example in Supplementary Figure S5). Similarly, the population set should consist of a distinct representation of the gene pool of a species. For example, it is problematic (but see Prunier *et al.*, 2012) to include different evolutionary lineages (Bragg *et al.*, 2015). Finally, marker density in relation to linkage disequilibrium will decisively affect the chances of confirming associations in an independent data set. If both linkage and marker density are low, the chances of finding similar signatures of selection in the original and independent data set might be minimal at the level of single SNP loci. In such a case, it would be wiser to sequence complete candidate genes. It has been shown in various organisms that adaptive convergence—adaptation to the same environmental conditions through the same genetic mechanisms—is often much higher among genes than among SNPs (e.g., Tenaillon *et al.*, 2012).

In conclusion, owing to complex local patterns of environment and genetic backgrounds, our study demonstrates a case where the putative adaptive role (i.e., environmental association) of many candidate SNPs could not be confirmed in an enlarged and independent set of populations. Still, we identified 11 genes (31%) for which at least one of the two applied statistical approaches confirmed the allele–environment association. This adds strong evidence that these genes have a significant role in large-scale adaptation to the abiotic environment, whereas most other genes reflect very local adaptation to a highly heterogeneous alpine environment.

#### DATA ARCHIVING

Raw data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.h9022>.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGEMENTS

We thank Alex Bösch, Elvira Schnyder, Jolanda Zimmermann, Deborah Zulliger and Nora Aellen for sampling; the Genetic Diversity Centre Zürich (GDC) for experimental support; Niklaus Zimmermann and Achilleas Psomas for making environmental data available; LGC genomics and Jane McDougall for KASP genotyping; Dorena Nagel for advice on data analysis; and two anonymous reviewers for valuable comments on a previous version of this manuscript. This study was funded by the Swiss National Science Foundation (project CRSI33\_127155) to AW, KKS and RH.

- Bianchi MW, Damerval C, Vartanian N (2002). Identification of proteins regulated by cross-talk between drought and hormone pathways in *Arabidopsis* wild-type and auxin-insensitive mutants, *axr1* and *axf2*. *Funct Plant Biol* **29**: 55–61.
- Bragg JG, Supple MA, Andrew RL, Borevitz JO (2015). Genetic variation across landscapes: insights and applications. *New Phytol* **207**: 953–967.
- Buehler D, Holderegger R, Brodbeck S, Schnyder E, Gugerli F (2014). Validation of outlier loci through replication in independent data sets: a test on *Arabis alpina*. *Ecol Evol* **4**: 4296–4306.
- Čalić I, Bussotti F, Martínez-García P, Neale D (2015). Recent landscape genomics studies in forest trees—what can we believe? *Tree Genet Genomes* **12**: 1–7.

- Clauss MJ, Koch MA (2006). Poorly known relatives of *Arabidopsis thaliana*. *Trends Plant Sci* **11**: 449–459.
- de Villemereuil P, Frichot E, Bazin E, François O, Gaggiotti OE (2014). Genome scan methods against more complex models: when and how much should we trust them? *Mol Ecol* **23**: 2006–2019.
- Farooq S, Farooq EA (2001). Co-existence of salt and drought tolerance in Triticeae. *Hereditas* **135**: 205–210.
- Feulner PGD, Chain FJJ, Panchal M, Huang Y, Eizaguirre C, Kalbe M *et al.* (2015). Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genet* **11**: e1004966.
- Fischer MC, Rellstab C, Tedder A, Zoller S, Gugerli F, Shimizu KK *et al.* (2013). Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Mol Ecol* **22**: 5594–5607.
- François O, Martins H, Caye K, Schoville SD (2016). Controlling false discoveries in genome scans for selection. *Mol Ecol* **25**: 454–469.
- Frichot E, Schoville SD, Bouchardeau G, François O (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol* **30**: 1687–1699.
- Godé C, Decombeix I, Kostecka A, Wasowicz P, Pauwels M, Courseaux A *et al.* (2012). Nuclear microsatellite loci for *Arabidopsis halleri* (Brassicaceae), a model species to study plant adaptation to heavy metals. *Am J Bot* **99**: e49–e52.
- Goslee SC, Urban DL (2007). The ecodist package for dissimilarity-based analysis of ecological data. *J Stat Softw* **22**: 1–19.
- Guillot G, Rousset F (2013). Dismantling the Mantel tests. *Methods Ecol Evol* **4**: 336–344.
- Hoekstra HE, Hirschmann RJ, Bunday RA, Insel PA, Crossland JP (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**: 101–104.
- Hohenlohe PA, Phillips PC, Cresko WA (2010). Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *Int J Plant Sci* **171**: 1059–1071.
- Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J *et al.* (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**: 55–61.
- Kawecki TJ, Ebert D (2004). Conceptual issues in local adaptation. *Ecol Lett* **7**: 1225–1241.
- Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A *et al.* (2011). PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* **6**: e15925.
- Korves TM, Schmid KJ, Caicedo AL, Mays C, Stinchcombe JR, Purugganan MD *et al.* (2007). Fitness effects associated with the major flowering time gene FRIGIDA in *Arabidopsis thaliana* in the field. *Am Nat* **169**: E141–E157.
- Lacombe B, Becker D, Hedrich R, DeSalle R, Hollmann M, Kwak JM *et al.* (2001). The identity of plant glutamate receptors. *Science* **292**: 1486–1487.
- Lotterhos KE, Whitlock MC (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol* **24**: 1031–1046.
- Mousavi SAR, Chauvin A, Pascaud F, Kellenberger S, Farmer EE (2013). Glutamate receptor-like genes mediate leaf-to-leaf wound signalling. *Nature* **500**: 422–425.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009). Divergent selection and heterogeneous genomic divergence. *Mol Ecol* **18**: 375–402.
- Parry G, Ward S, Cernac A, Dharmasiri S, Estelle M (2006). The *Arabidopsis* suppressor of auxin resistance proteins are nucleoporins with an important role in hormone signaling and development. *Plant Cell* **18**: 1590–1603.
- Pauwels M, Frérot H, Bonnin I, Saumitou-Laprade P (2006). A broad-scale analysis of population differentiation for Zn tolerance in an emerging model species for tolerance study: *Arabidopsis halleri* (Brassicaceae). *J Evol Biol* **19**: 1838–1850.
- Poncet BN, Herrmann D, Gugerli F, Taberlet P, Holderegger R, Gielly L *et al.* (2010). Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Mol Ecol* **19**: 2896–2907.
- Pritchard JK, Di Rienzo A (2010). Adaptation—not by sweeps alone. *Nat Rev Genet* **11**: 665–667.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Prunier J, Gerardi S, Laroche J, Beaulieu J, Bousquet J (2012). Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. *Mol Ecol* **21**: 4270–4286.
- Rellstab C, Gugerli F, Eckert A, Hancock A, Holderegger R (2015). A practical guide to environmental association analysis in landscape genomics. *Mol Ecol* **24**: 4348–4370.
- Rellstab C, Zoller S, Tedder A, Gugerli F, Fischer MC (2013). Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. *PLoS One* **8**: e80422.
- Rousset F (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour* **8**: 103–106.
- Schmidt PS, Serrao EA, Pearson GA, Riginos C, Rawson PD, Hilbish TJ *et al.* (2008). Ecological genetics in the north Atlantic: environmental gradients and adaptation at specific loci. *Ecology* **89**: S91–S107.
- Storey JD, Tibshirani R (2003). Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* **100**: 9440–9445.
- Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD *et al.* (2012). The molecular diversity of adaptive convergence. *Science* **335**: 457–461.
- Tiffin P, Ross-Ibarra J (2014). Advances and limits of using population genetics to understand local adaptation. *Trends Ecol Evol* **29**: 673–680.



- Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010). Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat Genet* **42**: 260–263.
- Volkert K, Debast S, Voll LM, Voll H, Schiessl I, Hofmann J *et al.* (2014). Loss of the two major leaf isoforms of sucrose-phosphate synthase in *Arabidopsis thaliana* limits sucrose synthesis and nocturnal starch degradation but does not alter carbon partitioning during photosynthesis. *J Exp Bot* **65**: 5217–5229.
- Weigel D, Nordborg M (2005). Natural variation in *Arabidopsis*. How do we find the causal genes? *Plant Physiol* **138**: 567–568.
- Zhao J, Li P, Motes CM, Park S, Hirschi KD (2015). CHX14 is a plasma membrane K<sup>+</sup>-efflux transporter that regulates K<sup>+</sup> redistribution in *Arabidopsis thaliana*. *Plant Cell Environ* **38**: 223–2238.
- Zimmermann NE, Kienast F (1999). Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *J Veg Sci* **10**: 469–482.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)