

# An outlier locus relevant in habitat-mediated selection in an alpine plant across independent regional replicates

**Journal Article****Author(s):**

Buehler, Dominique; Poncet, Bénédicte; Holderegger, Rolf; Manel, Stéphanie; Taberlet, Pierre; Gugerli, Felix

**Publication date:**

2013-03

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000064291>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

**Originally published in:**

Evolutionary Ecology 27(2), <https://doi.org/10.1007/s10682-012-9597-8>

## An outlier locus relevant in habitat-mediated selection in an alpine plant across independent regional replicates

Dominique Buehler · Bénédicte N. Poncet · Rolf Holderegger ·  
Stéphanie Manel · Pierre Taberlet · Felix Gugerli

Received: 6 July 2011 / Accepted: 20 July 2012 / Published online: 17 August 2012  
© Springer Science+Business Media B.V. 2012

**Abstract** Habitat types can induce genetic responses in species and may drive adaptive differentiation and evolutionary divergence of populations. In this study, we aimed at detecting loci indicative of adaptation for different habitat types in the alpine plant *Arabis alpina*. We used a dataset consisting of *A. alpina* plants collected in scree, nutrient-rich and moist habitat types in two independent regional replicates of the European Alps (the Swiss and French Alps). Genome scans resulting in 825 amplified fragment length polymorphisms (AFLPs) followed by outlier analysis, i.e. looking for excessive differentiation between habitat types, after accounting for heterozygosity and population structure, was used to detect loci under divergent selection for habitat type within and across the alpine regions. The outlier analyses resulted in the detection of a consistent single outlier locus, which showed a higher fragment frequency in moist compared to the other habitat types in both alpine regions. In addition, a posteriori tests for hierarchical population structuring in the dataset did not detect signals confounding selection at this locus (i.e. signals of regional population structure). Thus, we consider this locus indicative of habitat-mediated selection, and we subsequently sequence-characterized and compared it to the *Arabidopsis* genome. The sequence was found to be a putative homologue to the SIT4 phosphatase-associated family protein. The detection of this locus in two alpine regions and the availability of its

**Electronic supplementary material** The online version of this article (doi:[10.1007/s10682-012-9597-8](https://doi.org/10.1007/s10682-012-9597-8)) contains supplementary material, which is available to authorized users.

D. Buehler · R. Holderegger · F. Gugerli (✉)  
WSL Swiss Federal Research Institute, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland  
e-mail: felix.gugerli@wsl.ch

D. Buehler · R. Holderegger  
Department of Environmental Sciences, ETH Zürich, Universitätsstrasse 16, 8092 Zürich, Switzerland

B. N. Poncet · S. Manel · P. Taberlet  
Laboratoire d'Écologie Alpine (LECA), CNRS UMR 5553, Université Joseph Fourier, BP 53,  
2233 Rue de la Piscine, 38041 Grenoble cedex 9, France

S. Manel  
UMR 151 UP/IRD, Laboratoire Population–Environnement Développement, Université Aix-Marseille,  
3, place Victor Hugo, 13331 Marseille cedex 3, France

genome sequence make this locus a strong candidate worth further exploration in the habitat-mediated selection and genetic adaptation of natural populations in the alpine plant *A. alpina*.

**Keywords** Adaptation · *Arabis alpina* · Genome scan · Habitat type · Natural selection · Outlier analysis

## Introduction

The genetic response of a species to its local environment and to different habitat types is of great importance in evolution. Once adapted to different habitat types, populations of a species may experience reduced gene flow. For instance, this could be a consequence of limited pollen movement resulting from phenological shifts and subsequently weak gene flow that is unable to overcome local adaptation (Rieseberg and Willis 2007; Sobel et al. 2009). This process may ultimately cause the accumulation of genetic differences and potentially result in evolutionary divergence of ecotypes (Rieseberg and Willis 2007). Thus, habitat types have been recognized as important drivers of adaptive differentiation and genomic population divergence. For example, the marine gastropod *Littorina saxatilis* has been shown to harbour particular loci which are under divergent selection for contrasting habitat types along coastal gradients (Galindo et al. 2011; Wilding et al. 2001). Likewise, strong adaptation to soil type has been detected in *Arabidopsis lyrata* by mapping candidate mutations showing serpentine soil-mediated selection (Turner et al. 2010). However, especially in natural populations of non-model organisms, it remains a major challenge to identify genes and genomic regions underlying adaptive traits (Alonso-Blanco et al. 2009), such as those involved in adaptation to habitat types.

Genome scans followed by outlier analysis are seen as a major method to improve our understanding of the genetic basis of adaptation (Sgrò et al. 2011; Storz 2005). Outlier analysis is of particular use in non-model organisms since a priori information on the genomic background of the study species is not necessary (Storz 2005; Vasemägi and Primmer 2005). Instead, hundreds to thousands of usually anonymous genetic markers are genotyped for each sampled individual (Beaumont and Nichols 1996), and loci that show a higher or lower differentiation than expected under neutrality are detected as outlier loci indicative of divergent or balancing selection, respectively (Luikart et al. 2003; Storz 2005; Vasemägi and Primmer 2005). These outlier loci are either directly located in adaptive genes or, more likely, they are only linked to genes and genomic regions under selection (Nosil et al. 2009). Thus far, a large number of outlier loci, largely those indicative of divergent selection, have been identified across a wide range of organisms and study questions (reviewed in Holderegger et al. 2008; Vasemägi and Primmer 2005). However, the sequences underlying outlier loci detected on the basis of anonymous markers such as amplified fragment length polymorphisms (AFLPs) have rarely been characterized (but see Minder and Widmer 2008; Wood et al. 2008), although this is an important next step in identifying genes under adaptive selection in non-model species.

The adaptive signal of outlier loci (i.e. especially high population divergence) should be caused by environmental selection. However, historical and demographical processes may induce substantial spatial genetic structure (e.g. bottlenecks, range expansion, admixture zones) and can thus be confounded with the signal of adaptation at outlier loci (Robertson 1975; Schlötterer 2003). Outlier analysis tends to underestimate these demographical effects, which increases the number of false positives (Excoffier et al. 2009; Hofer et al. 2009).

Large-scale population genetic structure is found in alpine ecosystems since populations diverged in isolation during glacial periods and subsequently came into secondary contact through re-expansion (Alvarez et al. 2009; Körner 2003; Schönswetter et al. 2005; Thiel-Egenter et al. 2011). Under such circumstances, limiting confounding effects on signals of divergent selection and detecting real outlier loci remains challenging, but can be achieved by accounting for hierarchical population structure (Excoffier et al. 2009). Moreover, studying independent regional replicates may reduce confounding signals and restrict confidence to outlier loci detected across regions (e.g. Manel et al. 2010b; Poncet et al. 2010). This is based on the reasoning that neutral historical or demographical processes would not create the same genetic pattern across independent replicates. Thus, replicated regional sampling as in the present study has been advocated (Holderegger et al. 2010) to help eliminating potentially false positives, i.e. giving more confidence in those loci identified as outliers.

We studied the alpine plant *Arabis alpina* L. (Brassicaceae), which occurs in distinct habitat types, namely (1) scree, (2) nutrient-rich and (3) moist habitats, across two alpine regions in the Swiss and French Alps. A previous study used the same genome scan dataset to search for loci involved in the general response of *A. alpina* to different environmental variables related to temperature, precipitation and topography with regression analysis (Poncet et al. 2010). In turn, our objective in the present study was to analyze the dataset using outlier analysis to detect specific genomic regions involved in genetic adaptation to different habitat types as classified a priori. *A. alpina* has recently become a thoroughly studied plant species for ecological and population genetic aspects (Ansell et al. 2008; Assefa et al. 2007; Buehler et al. 2012; Manel et al. 2010b; Poncet et al. 2010; Tedder et al. 2011; Wang et al. 2009), which makes it a suitable organism to inquire about the genetic basis of local adaptation. Moreover, incorporating adaptive responses in species distribution models will help improving scenarios of future vegetation under the perspective of global change (Jay et al. 2012 and references therein).

The strategy of this study was to detect outlier loci by grouping individuals from all sampling locations according to the three habitat types in which they were collected. We then performed outlier analyses using these three habitat type groupings. Our expectation was to detect outlier loci which occur both within and across the alpine regions and to detect consistent patterns of allele frequencies among habitat types at these loci in both regions. Furthermore, we tested for patterns of hierarchical population structure to avoid loci potentially affected by demographic processes. Finally, we characterized the molecular basis of the most consistent outlier locus and compared it to the whole-genome information of the model species *Arabidopsis thaliana* and its relatives.

## Materials and methods

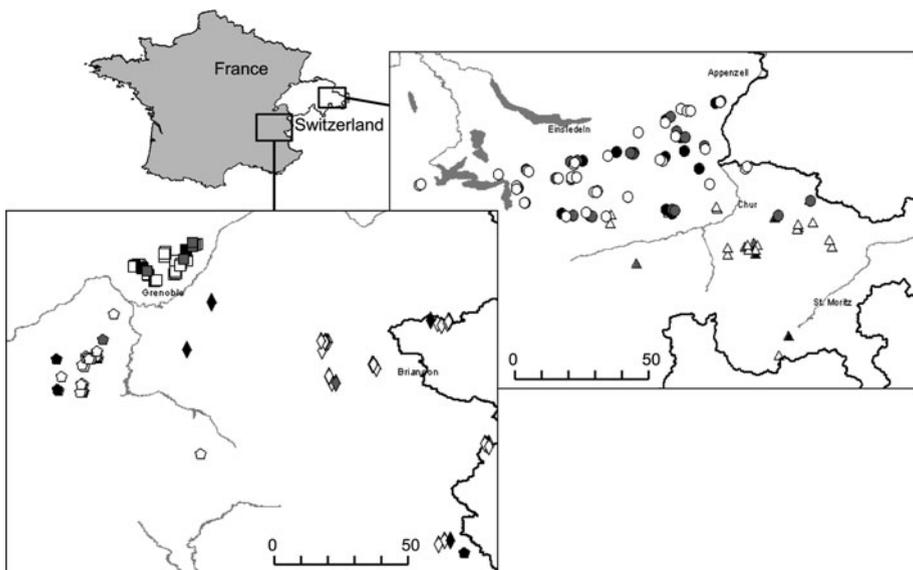
### Study species, study area and AFLP genotyping

The alpine rock-cress *A. alpina* L. (Brassicaceae) is a perennial rosette herb. It has a broad distribution ranging from the amphi-atlantic region to the European mountain system (Koch et al. 2006). In the European Alps, *A. alpina* grows between 400 m and 3,200 m a.s.l. on calcareous substrates in various open habitats (Titz 1971). It reproduces sexually, presumably with a substantial rate of inbreeding (Ansell et al. 2008; Buehler et al. 2012), or vegetatively via stolons (Schultze-Motel 1986). *A. alpina* is a wild relative of

*A. thaliana* for which whole-genome sequence information is available (The Arabidopsis Initiative 2000).

We used an existing AFLP dataset on *A. alpina* (Herrmann et al. 2010; Poncet et al. 2010). Samples from 192 sites in two regions (i.e. Swiss and French Alps) were collected at elevations ranging from 440 m to 3,133 m a.s.l. in summer 2006 (Fig. 1). Samples originated from three distinct habitat types: scree, nutrient-rich and moist (Table 1). Scree habitats are found in large scree fields that are dynamic, dry or with irregular water availability and low humus content. The nutrient-rich habitat is characterized by high nutrient content and is naturally found in the alpine ecosystem under exposed rocks and along ridges frequently visited by wild animals, recognized by unusually rich vegetation and respective indicator species. Nutrient-rich habitats also occur at anthropogenically influenced sites such as alpine pastures near cattle farms. The moist habitat is defined by high water availability and high humus content, it is mostly found along small streams. At each sampling location, fresh leaf material was collected from three to nine individuals and immediately dried in silica gel.

Initially, 2386 AFLP loci were genotyped with 19 primer/enzyme combinations (Herrmann et al. 2010; Poncet et al. 2010). AFLP loci were automatically selected according to the stringent procedure implemented in SCANFLP (Herrmann et al. 2010). AFLP loci with low reproducibility or minor polymorphisms (i.e. <3 individuals with a different presence/absence score than all other samples) were discarded. The final dataset consisted of 825 polymorphic loci in 634 individuals. Poncet et al. (2010) detected linkage disequilibrium for only 3.5 % of locus pairs.



**Fig. 1** Map showing the 192 sampling locations of *A. alpina* in the two study regions, the French and the Swiss Alps. Different symbols represent STRUCTURE clusters (Falush et al. 2003, 2007; Pritchard et al. 2000) of the sampling locations within the Swiss and French Alps, as calculated from our dataset. Habitat types are given by different shading (scree: black; nutrient-rich: grey; moist: white)

**Table 1** Number of AFLP loci ( $L$ ) and individuals of *A. alpina* collected for each habitat type (scree, nutrient-rich and moist) used in three analyses of the Swiss Alps, the French Alps and the cumulative dataset (i.e. the Swiss and French Alps combined)

	$L$	Number of individuals		
		Scree	Nutrient-rich	Moist
Swiss Alps	443	177	96	71
French Alps	503	236	30	36
Cumulative dataset	523	413	126	107

### Outlier detection for habitat type

To detect outlier loci, we searched for loci that exhibited higher genetic differentiation (i.e. divergent selection) among the three habitat types than expected under neutrality with the program `DFDIST` (Beaumont and Balding 2004; Beaumont and Nichols 1996). `DFDIST` uses coalescent simulation under Wright's (1951) symmetrical island model. A neutral distribution of  $F_{ST}$  values, i.e. genetic differentiation between habitat types using the method of Weir and Cockerham (1984), is generated conditional on expected heterozygosity ( $H_e$ ) and based on thousands of simulated loci with a trimmed mean  $F_{ST}$  identical to the mean empirical  $F_{ST}$ . This null distribution is used to separate outlier loci from neutral loci based on confidence intervals. For each analysis, `DFDIST` was configured to generate 50,000 loci and  $N_e$  was set to 1,000. We designated those loci as outliers which had an observed  $F_{ST}$  value higher than the upper confidence limit ( $P = 0.05$ ).

We first performed the above `DFDIST` analysis for habitat types separately in the Swiss and French Alps, and secondly in the cumulative dataset combining the samples from the two alpine regions. The number of loci used as well as the number of individuals per habitat type in the Swiss and French Alps and in total are given in Table 1. We considered loci to be consistent outlier loci for habitat type if they were identified in both regions and in the cumulative analysis.

### Identification of hierarchical population structure

In order to detect hierarchical population genetic structure, hereafter referred as population structure, we first applied the Bayesian clustering in `STRUCTURE` v2.3.1 modified for dominant data (Falush et al. 2003, 2007; Pritchard et al. 2000). A previous analysis by Poncet et al. (2010) detected two clusters in the whole dataset ( $F_{ST} = 0.1652$ ), namely the Swiss and the French Alps. Based on this observation, we considered the Swiss Alps and the French Alps to represent two independent regional replicates. We thus performed `STRUCTURE` analyses separately for each of the two alpine regions to detect population structure within these regions. Each sampling location per region was associated with any of the large-scale genetic clusters detected in a previous study with an independent dataset (Alvarez et al. 2009). In case a sampling location occurred in an area of mixed cluster membership in Alvarez et al. (2009), all individuals of such a location were given a `LOCPRIOR` that matched the cluster with the highest assignment probability in the adjacent sampling locations of Alvarez et al. (2009). We then used this prior grouping within regions for the `LOCPRIOR` function in `STRUCTURE` (Hubisz et al. 2009). Individuals were then clustered into  $K$  discrete clusters ( $K = 1-6$ ) with the admixture model. We performed three independent runs for each  $K$ , with a burn-in period of 50,000 cycles and

50,000 Markov Chain Monte Carlo replications, which was sufficient to give convergent results. Using no location prior or taking the coordinates of each sampling location instead of grouping according to Alvarez et al. (2009) gave very similar results (data not shown). We determined the optimal value of  $K$  by visually inspecting the change in maximum likelihood probability and increased admixture as opposed to new subgroups (Fig. S1). Second, analysis of molecular variance (AMOVA) as implemented in ARLEQUIN v3.1 (Excoffier et al. 2005) was carried out to test for genetic differentiation among STRUCTURE groups using 1 000 permutations. AMOVA computes  $\phi_{ST}$ , an analog of  $F_{ST}$  statistics (Excoffier et al. 1992).

#### Identification of population structure in the outlier loci detected

To test whether the outlier loci we detected reflected population structure, we calculated locus-specific  $\phi_{ST}$ s (Excoffier et al. 2005) for all 825 polymorphic loci. We determined three separate  $\phi_{ST}$ -values per locus, namely (1) between the Swiss Alps and the French Alps to identify loci with large-scale population structure, (2) between STRUCTURE groups in the Swiss Alps and (3) between STRUCTURE groups in the French Alps. The latter two should detect loci with small-scale population structure. We used results from our STRUCTURE analyses, classifying all individuals per location to a respective cluster according to a majority rule. Significance was estimated with 1,000 permutations. Corresponding threshold values for  $\phi_{ST}$  were determined from  $F_{SC}$  bootstrap percentile values at  $P = 0.05$  in ARLEQUIN. Loci which were detected as significant were designated as exhibiting population structure. Note that outlier loci showing such structure are not necessarily non-adaptive, as variation in population structure may correlate with a regional or large-scale environmental gradient (Holderegger et al. 2008).

#### Sequencing of outlier locus

Only one locus was identified as a consistent outlier locus across all outlier analyses and was also not affected by population structure (see Results). For this locus EM74.7, we generated AFLP genotypes using the procedures in Herrmann et al. (2010) for eight samples: four samples that lacked the corresponding AFLP band and four samples with the AFLP band present. Of the eight samples (three from the French Alps, five from the Swiss Alps), three were collected in scree, two in nutrient-rich and three in moist habitat types. Genomic DNA was digested using *EcoRI/MseI* and the selective PCR was done using primers *EcoRI* 5'-GACTGCGTACCAATTC with selective bases ATC and *MseI* 5'-GATGAGTCCTGAGTAA with selective bases CAC.

The AFLP locus EM74.7 was then isolated using the procedure of Roden et al. (2009). Electrophoresis of AFLP bands was done on Spreadex<sup>®</sup> EL 500 mini gels (Elchrom Scientific, Cham, Switzerland) in 30 mM Tris–acetate EDTA (TAE) buffer using 10  $\mu$ L of selective PCR product and M3 Marker (Elchrom Scientific, Cham, Switzerland) as size standard. After pre-electrophoresis for 10 min at 50 V, samples were loaded and run for 82 min at 120 V. The gel was stained with SYBR GOLD (Clare Chemical Research, Dolores, CO, USA) for 30 min, and the bands were viewed in an Epi Chemi II Darkroom (UVP Laboratory Products, Upland, CA, USA). We excised locus EM74.7 from samples with the amplified band using a cylinder of 1 mm diameter. As a control, two longer fragments were also excised from the gel. Each gel cut was eluted in 25  $\mu$ L deionized water for 24 h at 4 °C. We repeated this procedure and excised locus EM74.7 from a second Spreadex<sup>®</sup> gel.

The excised DNA fragments were amplified using primers designed for the adaptor sequences (*EcoRI* + 0 and *MseI* + 0) as described in Roden et al. (2009). The PCR products were purified using the MinElute Kit (QIAGEN, Hilden, Germany), and sequencing reactions were performed in both directions with BigDye<sup>®</sup> Terminator Cycle Sequencing Kit v1.1 (Applied Biosystems, Foster City, CA, USA). Sequences were determined on an ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Alignment of sequences was done using CLC SEQUENCER VIEWER (CLC bio, Aarhus, Denmark), and the obtained consensus sequence was BLASTed against the nucleotide database of GenBank (Altschul et al. 1990).

## Results

### Outlier analysis

The DFDIST analyses identified 34 outlier loci (7.7 % of all loci) in the Swiss Alps, 37 outlier loci (7.4 %) in the French Alps and 32 outlier loci (6.1 %) in the cumulative dataset of the Swiss and French Alps combined (Fig. 2). Thus, the percentage of outlier loci detected was consistent with previous genome scan studies (Bonin et al. 2006). Two loci, EM74.7 and PM179.6, were the only outlier loci detected in both alpine regions (Figs. 2, 3a). Locus EM74.7 was also detected as an outlier locus in the cumulative dataset. Furthermore, locus EM74.7 consistently had a higher frequency of fragment presence in moist habitats in all analyses (Fig. 4).

### Population structure

Based on results of STRUCTURE, we observed two clusters in the Swiss Alps, namely the southeastern Swiss Alps and the central/northeastern Swiss Alps ( $\phi_{ST} = 0.1318$ ,  $P < 0.001$ ). These two clusters only partially matched the groupings created by Poncet et al. (2010) owing to our dataset being reduced to Swiss samples only. In the French Alps, three clusters were identified, namely the two mountain massifs Vercors and Chartreuse and the region Brionçonnais ( $\phi_{ST} = 0.2812$ ,  $P < 0.0001$ ; Figs. 1, S1). There was no coincidence of identified clusters with habitat types.

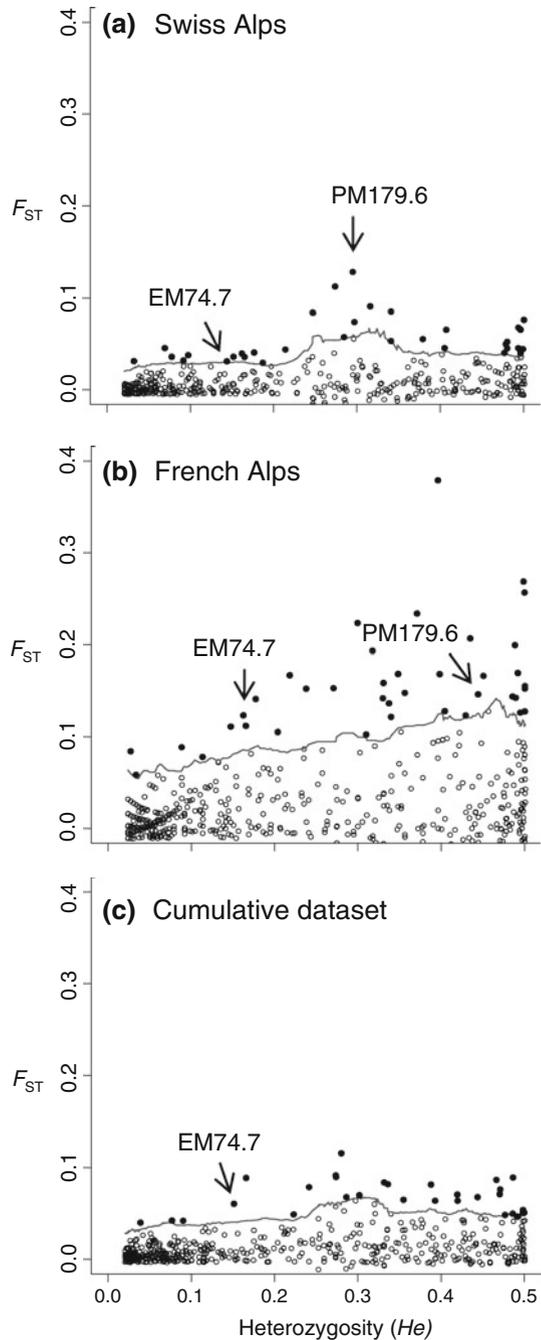
We found significant large-scale population structure for a total of 113 loci (13.7 % of all loci; corresponding threshold value  $\phi_{ST} = 0.25$ ). This reduced the number of confident outlier loci in each DFDIST analysis (Fig. 3b), but neither affected locus EM74.7 nor locus PM179.6.

In the next step, we detected 104 loci (12.6 % of all loci; corresponding threshold value  $\phi_{ST} = 0.15$ ) with small-scale population structure among the STRUCTURE groups within the Swiss Alps. This reduced the number of confident outlier loci in the Swiss Alps, but also did not affect loci EM74.7 and PM179.6 (Fig. 3c). Finally, we detected 110 loci (13.3 % of all loci; corresponding threshold value  $\phi_{ST} = 0.30$ ) with small-scale population structure between the STRUCTURE groups in the French Alps, reducing the number of confident outlier loci in the French Alps, but again not affecting loci EM74.7 and PM179.6 (Fig. 3c).

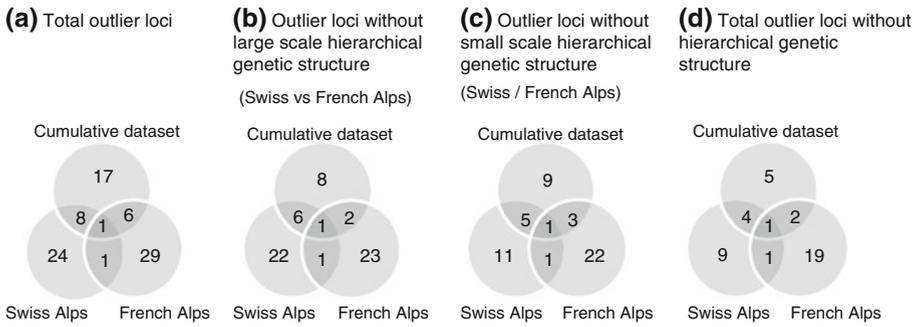
### Outlier sequence and its localization in the *Arabidopsis* genome

We sequenced 41 bp (78 bp including selective bases and both adaptors) of locus EM74.7 (GenBank accession no. HM594277.1), which was the only AFLP locus consistently

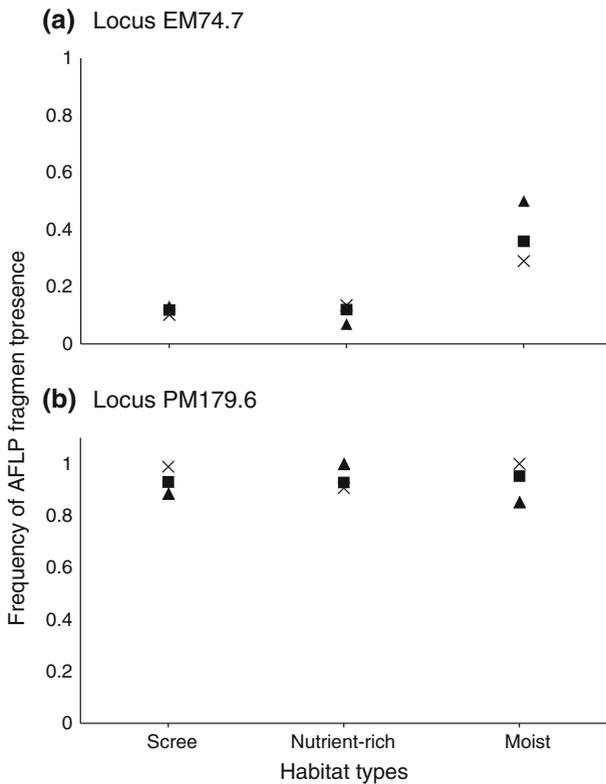
**Fig. 2** Plots of  $F_{ST}$  analyses (Beaumont and Balding 2004; Beaumont and Nichols 1996) to detect outlier loci that are more strongly differentiated among habitat types than expected under neutrality in *A. alpina*. The distribution of  $F_{ST}$  values is shown as a function of heterozygosity ( $H_e$ ) for (a) the Swiss Alps, (b) the French Alps, and (c) the cumulative dataset (i.e. the Swiss and French Alps combined). The *solid line* depicts the 95 % confidence interval, and each dot represents a single AFLP locus. *Filled dots* above the *solid line* indicate outlier loci indicative of divergent selection, and *open circles* below the *solid line* designate neutral loci. Marked with an *arrow* are the outlier loci EM74.7 and PM179.6



detected as an outlier locus in all analyses and not showing locus-specific population structure. In a nucleotide BLAST search, locus EM74.7 gave four significant hits (E-value < 0.16; note that short sequences often have relatively high E-values). The



**Fig. 3** Venn diagrams showing the effect of significant hierarchical population structure on the number of outlier loci among habitat types in *A. alpina*. Each Venn diagram shows the overlap of outlier loci detected in the Swiss Alps, the French Alps, and the cumulative dataset (i.e. the Swiss and French Alps combined). (a) Total number of outlier loci found in all three outlier analyses. (b) Number of outlier loci showing no hierarchical population structure on large spatial scale, between the Swiss and French Alps. (c) Number of outlier loci showing no hierarchical population structure at small spatial scale among STRUCTURE groups within the Swiss and French Alps. (d) Total number of outlier loci showing no hierarchical population structure



**Fig. 4** Frequency of presence of (a) locus EM74.7 and (b) locus PM179.6 in *A. alpina* across the three habitat types scree, nutrient-rich and moist. Crosses: Swiss Alps; Triangles: French Alps; Squares: Cumulative dataset

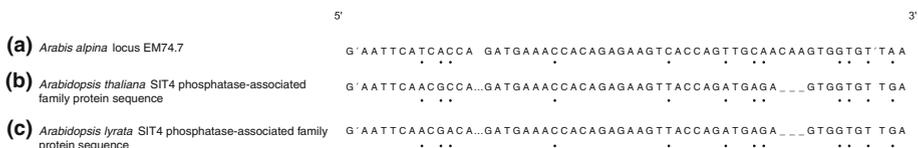
sequence of locus EM74.7 matched with the SIT4 phosphatase-associated family protein in *A. lyrata* (GenBank accession no. XM\_002890826.1, max. score 42.1, max. identity = 96 %). In *A. thaliana*, the outlier sequence also matched with the SIT4 phosphatase-associated family protein (At1g30470; GenBank accession no. NM\_102783.4, max. score 42.1, max. identity = 96 %) as well as with a full length cDNA sequence associated with gene At1g30470 (GenBank accession no. BX818019.1, max. score 42.1, max. identity = 96 %), and a complete BAC sequence on chromosome 1 (GenBank accession no. AC009917.2, max. score 42.1, max. identity = 96 %). By aligning these sequences, regions of high conservation were revealed among species. Figure 5 illustrates the alignment of the short sequence of locus EM74.7 in *A. alpina* and the SIT4 phosphatase-associated family proteins in *A. lyrata* and *A. thaliana*.

## Discussion

Species inhabiting different habitat types may be exposed to divergent selection pressures, which could lead to adaptive population differentiation. Genome scans followed by outlier analysis are a feasible method to detect genes or (more likely) genomic regions under selection for habitat types. This method is applicable to non-model organisms which lack genomic background information. However, sequence-characterizing outlier loci is a step rarely done in outlier studies using non-model organisms. In this study, our goal was to detect habitat-mediated selection in the alpine plant *A. alpina* by comparing samples from different habitat types across two independent regions of the European Alps. In a suite of stringent analyses, we found one locus as a consistent outlier locus for downstream analysis, whose sequence matched to a coding region in *Arabidopsis* genomes. In the following, we interpret the environmental link of the consistent outlier locus detected and discuss the stringency and limitations of our study.

### Outlier loci

The outlier analyses performed in this study revealed that divergent selection affected 6–7.7 % of loci among all analyses. Almost all of these outlier loci were limited to either the Swiss or French Alps and were not replicated across the two alpine regions. This may be a result of (1) most outlier loci showing only weak selection, (2) selection mainly acting on a local scale or (3) the detection of false positive outlier loci (Minder and Widmer 2008). On the other hand, outlier loci which are detected as outlier loci in several independent regions or environmental gradients may be considered to be under replicated divergence (Nosil et al. 2009; Schmidt et al. 2008). Therefore, we applied such a strict



**Fig. 5** Alignment of sequences of locus EM74.7 in (a) *A. alpina* and the putative homologue sequences of the SIT4 phosphatase-associated family protein in (b) *A. thaliana* (GenBank accession no. NM\_102783.4) and (c) *A. lyrata* (GenBank accession no. XM\_002890826.1). Nucleotide differences in the sequences are marked with dots

criterion to our dataset and only retained those outlier loci detected in both alpine regions under study. The Swiss and French Alps were considered as independent since they were significantly differentiated ( $F_{ST} = 0.1652$ ,  $P < 0.0001$ ) in a STRUCTURE analysis of our dataset by Poncet et al. (2010). In addition, a phylogeographic analysis of *A. alpina* from the Alps in Ansell et al. (2008) showed populations in the two alpine regions as diverged lineages, which was confirmed by Alvarez et al. (2009).

By the approach of replicating outlier analyses across the alpine regions, we were left with two consistent outlier loci, EM74.7 and PM179.6. Similar studies using outlier analysis have also remained with a low number of outlier loci once replicates across regions or environmental gradients were considered (e.g. Miller et al. 2007; Nosil et al. 2009; Oetjen et al. 2010; Poncet et al. 2010). This could be a result of some potentially relevant outlier loci being discarded due to the stringency. Nevertheless, our consistent outlier loci show the strongest signals of selection and are not likely to be affected by population structure (see below). Therefore, they are best suited for downstream applications such as sequence characterization, reciprocal transplant experiments or expression studies (Holderegger et al. 2008).

Locus EM74.7 was considered to be a particularly strong candidate because this locus was replicated as an outlier across the alpine regions and also detected in a cumulative analysis. We considered such a pattern as pre-condition for a reliable outlier under the assumption that an allele under selection for a particular environmental response such as habitat type is potentially targeted by selection across all sampling sites (Verhoeven et al. 2008). Locus EM74.7 further showed a significantly higher frequency of fragment presence in moist as compared to the other two habitat types (Fig. 4a). Such a significant change in allele presence/absence among habitat types or along environmental gradients conforms to our expectations. Even though locus PM179.6 was also detected across both alpine regions, this locus was of less interest because it was not detected in the cumulative analysis because its fragment frequencies were not consistently correlated with habitat types in both regions (Fig. 4b). Therefore, we considered locus EM74.7 to be best suited for downstream analysis, while we cannot rule out that locus PM179.6 is also within or linked to a genomic region under selection.

### Habitat-mediated divergence

Previous studies using outlier analysis in habitat-mediated selection studies have identified particular loci responsible for local adaptation (Galindo et al. 2011; Keller et al. 2010; Shikano et al. 2010; Wilding et al. 2001). The classical example is the detection of outlier loci in *L. saxatilis* for ecotypes with different shell types occurring in distinct habitat types across coastal shores (Wilding et al. 2001). Recently, a genome scan using next-generation sequencing has revealed that functional annotations of contigs containing outlier loci for these *Littorina* ecotypes are coding for shell matrix and muscle proteins (Galindo et al. 2011).

In order to detect the underlying function of locus EM74.7, we, likewise, isolated and sequence-characterized this locus and BLASTed the sequence against the *Arabidopsis* genome. The sequence obtained matched the SIT4 phosphatase-associated family protein (SAPs) in *A. thaliana* and *A. lyrata*. This protein is potentially conserved in the Brassicaceae family since it was detected in *A. thaliana*, *A. lyrata*, and *A. alpina*, the latter of which is phylogenetically distantly related (see Beilstein et al. 2006), however the role of SAPs in plants is unknown. In *Saccharomyces cerevisiae*, SAPs interact with a catalytic subunit (SIT4) of a type 2A-related protein phosphatase (Luke et al. 1996). SIT4 functions

in the execution of the start phase in the cell cycle (Luke et al. 1996). During this phase, cells may exit the cell cycle as a response to nutrient limitations or commit to cell division. Experiments in yeast have revealed that SAPs are unable to function in the absence of SIT4 and vice versa (Luke et al. 1996). However, without additional information of the underlying function or the selection factors involved in *A. alpina*, the role of SAPs in habitat-mediated selection remains speculative. Moreover, it should be kept in mind that SAP might not itself be the target of selection, but only in linkage disequilibrium with the gene of adaptive relevance.

The selective force behind habitat-mediated selection in *A. alpina* is not likely as clear as in the case of shell shape in *Littorina*. Outlier analysis only tests for population divergence and does not directly determine the particular selection factor acting upon the outlier loci or the linked genomic region (Stinchcombe and Hoekstra 2007; Storz 2005). In another analysis of the same dataset, Poncet et al. (2010) used a regression analysis in an environmental association study to detect loci correlated to environmental variables extracted from topo-climatic GIS layers of the European Alps. They detected several loci which were common in both the Swiss and French Alps and were related to mean annual minimum temperature. Thus, using an environmental association analysis has the advantage that a putative selection force can be determined such as annual minimum temperature. However, the selective agent behind adaptation to scree, nutrient-rich and moist habitat types most likely integrates across many ecological factors and temporal dynamics which cannot be captured by annual means such as used by Poncet et al. (2010). Along the same line, environmental association studies often rely on environmental data of coarse resolution and which are interpolated on the basis of measurements from climatic stations (Gugerli et al. 2008). In turn, our sampling considered distinct habitat types with characteristic site conditions, which we see as an important consideration for outlier and environmental association studies in the future. When comparing our results with the findings of Poncet et al. (2010) and Manel et al. (2010b), we note that locus EM74.7 was not identified as a marker that associated to any of the environmental parameters tested by Poncet et al. (2010), whereas Manel et al. (2010b) found only weak associations of this marker at the local scale, notably to summer precipitation in two of the three areas of the French region (S. Manel, unpublished data). The latter suggests that water availability in a given site may indeed be considered a crucial driver of local adaptation. On the other hand, habitat-mediated adaptation may involve the interaction of multiple selection factors ranging from water use efficiency or anoxia tolerance to nitrogen uptake or competition (reviewed in Reich et al. 2003). In the serpentine-soil mediated selection described by Turner et al. (2010), the authors speculated that candidate single nucleotide polymorphisms detected in the genome of *A. lyrata* may result from deleterious mutations which cause non-adapted individuals to perish. Similarly, we could consider that non-adapted individuals (individuals with a mutation at locus EM74.7 or the corresponding linked gene) of *A. alpina* had a lower fitness in moist habitats and were ultimately unable to grow and reproduce. However, follow-up investigations should be conducted to test whether such fitness advantages exist under natural conditions.

### Population structure

To rigorously test the reliability of our outlier loci, we attempted to distinguish signals of selection from neutral signals of historical or demographical events causing population structure. Historical or demographical signals can mimic selection and lead to the detection of a large number of false positive outlier loci in outlier analysis (Excoffier et al. 2009).

Loci EM74.7 and PM179.6 in *A. alpina* were the only two AFLP markers detected as consistent outliers across two independent regional replicates. Therefore it is unlikely that the effects of neutral processes caused high population divergence among habitat types in both regions (Bonin et al. 2006; Campbell and Bernatchez 2004; Luikart et al. 2003; Nosil et al. 2009; Storz 2005; Vasemägi and Primmer 2005). As a second test, we a posteriori searched for substantial population structure at a small and a large spatial scale in identified outlier loci. This differs from recently developed methodologies for outlier detection, which a priori test for population structure (Excoffier et al. 2009). As expected, a large number of outlier loci showed confounding population structure at both scales. However, the strongest candidate, locus EM74.7, as well as locus PM179.6 were not affected by population structure. This suggests that neutral processes, as far as we could test, were not interfering with the signals of selection at these outlier loci. Nevertheless, it should be noted that loci showing strong population genetic structure are not necessarily neutral; so removing them from the list of outlier loci is only a precaution to avoid targeting at potentially false positives for downstream analyses, e.g. the functionality of a particular locus.

## Conclusion

Identifying loci linked to genomic regions potentially under selection is still part of the exploratory stage in the work to find genes underlying relevant adaptive genetic variation (Manel et al. 2010a; Reusch and Wood 2007). Although we have gone one step further than most previous investigations in this field by sequence-characterizing the outlier locus of interest, we are still limited by the functional proof of the evolutionary relevance of the genomic polymorphisms detected. In theory, the underlying genomic position of an outlier locus is not assumed to be the direct target of selection, and at this point we lack the experimental evidence to interpret the gene(s) to which this outlier locus is linked. In our dataset, Poncet et al. (2010) detected linkage disequilibrium for 3.5 % of the 825 loci tested. In the analysis of Poncet et al. (2010), a large number of loci associated to environmental variables showed linkage, potentially pinpointing a genomic region of interest. However, in our analysis we only detected one locus of interest, therefore, we cannot evaluate its linkage with other outliers or genes of ecological relevance. On the other hand, we also cannot exclude the possibility that the coding region to which our outlier locus matched indeed represents the “needle in the haystack” and is directly involved in habitat-mediated selection. It is more likely that the needle has not been found but that we are zooming in on the section of hay in which it is hidden. Therefore, we believe that in this study we found a candidate genomic region for further intensive evaluation once the full genome sequence of *A. alpina* is available, e.g. testing for adaptive responses in field trials and functional tests at the molecular level (Kawecki and Ebert 2004).

The detection of a putative outlier locus that is consistently related to habitat types offers the possibility to study adaptive genetic diversity across complex landscapes. Especially with the present concern of climate change, isolating adaptive genes will enable us to measure how beneficial gene variants are distributed and may spread among populations (Holderegger et al. 2010), revealing the adaptive potential of populations and their putative range changes. A next step is to detect the polymorphisms causing the genetic pattern at outlier loci and to confirm the linkage of identified outlier loci to known genes under selection. Final proof should come from testing the adaptive relevance of putative outlier loci in their natural environment. For instance, the allelic frequency of an outlier

locus found in a particular environmental setting could be verified across a set of independent natural samples. In conclusion, this study represents a first step towards understanding the molecular basis of habitat-mediated selection for the alpine plant *A. alpina*, and we consider locus EM74.7 indicative of a genomic region of high interest in this species' adaptation to habitat type worth further genomic exploration.

**Acknowledgments** We would like to thank Doris Herrmann, René Graf, Conny Thiel-Egenter, Annina Bürgi, Nathalie Baumgartner, Fabio Rimensberger, Rolland Douzet, Serge Aubert, Ludovic Gielly, Delphine Rioux and Claire Redjadj for help in sampling and AFLP genotyping, Sabine Brodbeck for sequence characterization, György Sipos and Christoph Sperisen for advice on the functioning of SIT4 phosphatase-associated family proteins, and Sarah Bryner, Debbie Zulliger and several anonymous reviewers for their valuable comments on the manuscript. Funding was provided by the CCES-BIO-CHANGE project of the ETH domain. S. M. was supported by the Institut Universitaire de France.

## References

- Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M (2009) What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell* 21:1877–1896
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Alvarez N, Thiel-Egenter C, Tribsch A, Holderegger R, Manel S, Schönswetter P, Taberlet P, Brodbeck S, Gaudeul M, Gielly L, Küpfer P, Mansion G, Negrini R, Paun O, Pellecchia M, Rioux D, Schüpfer F, van Loo M, Winkler M, Gugerli F, IntraBioDiv Consortium (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecol Lett* 12:632–640
- Ansell SW, Grundmann M, Russell SJ, Schneider H, Vogel JC (2008) Genetic discontinuity, breeding-system change and population history of *Arabis alpina* in the Italian Peninsula and adjacent Alps. *Mol Ecol* 17:2245–2257
- Assefa A, Ehrlich D, Taberlet P, Nemomissa S, Brochmann C (2007) Pleistocene colonization of afro-alpine 'sky islands' by the arctic-alpine *Arabis alpina*. *Heredity* 99:133–142
- Beaumont M, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* 13:969–980
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc B* 263:1619–1626
- Beilstein MA, Al-Shehbaz A, Kellogg EA (2006) Brassicaceae phylogeny and trichome evolution. *Am J Bot* 93:607–619
- Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol Biol Evol* 23:773–783
- Buehler D, Graf R, Holderegger R, Gugerli F (2012) Contemporary gene flow and mating system of *Arabis alpina* in a Central European alpine landscape. *Ann Bot* 109:1359–1367
- Campbell D, Bernatchez L (2004) Genomic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol Biol Evol* 21:945–956
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50
- Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* 103:285–298
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7:574–578

- Galindo J, Grahame JW, Butlin RK (2011) An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. *J Evol Biol* 23:2004–2016
- Gugerli F, Englisch T, Niklfeld H, Tribsch A, Mirek Z, Ronikier M, Zimmermann NE, Holderegger R, Taberlet P, IntraBioDiv Consortium (2008) Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation—a project synopsis. *Perspect Plant Ecol Evol Syst* 10:259–281
- Herrmann D, Poncet BN, Manel S, Rioux D, Gielly L, Taberlet P, Gugerli F (2010) Selection criteria for scoring amplified fragment length polymorphisms (AFLPs) positively affect the reliability of population genetic parameter estimates. *Genome* 53:302–310
- Hofer T, Ray N, Wegmann D, Excoffier L (2009) Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Ann Hum Genet* 73:95–108
- Holderegger R, Herrmann D, Poncet B, Gugerli F, Thuiller W, Taberlet P, Gielly L, Rioux D, Brodbeck S, Aubert S, Manel S (2008) Land ahead: using genome scans to identify molecular markers of adaptive relevance. *Plant Ecol Divers* 1:273–283
- Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. *Trends Plant Sci* 15:675–683
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Res* 9:1322–1332
- Jay F, Manel S, Alvarez N, Durand E, Thuiller W, Holderegger R, Taberlet P, François O (2012) Forecasting changes in population genetic structure of Alpine plants in response to global warming. *Mol Ecol* 21:2354–2368
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecol Lett* 7:1225–1241
- Keller I, Taverna A, Seehausen O (2010) Evidence of neutral and adaptive genetic divergence between European trout populations sampled along altitudinal gradients. *Mol Ecol* 20:1888–1904
- Koch MA, Kiefer C, Ehrich D, Vogel J, Brochmann C, Mummenhoff K (2006) Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (Brassicaceae). *Mol Ecol* 15:825–839
- Körner C (2003) Alpine plant life: functional plant ecology of high mountain ecosystems. Springer, Berlin
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* 4:981–993
- Luke MM, Della Seta F, Di Como CJ, Sugimoto H, Kobayashi R, Arndt KT (1996) The SAPs, a new family of proteins, associate and function positively with the SIT4 phosphatase. *Mol Cell Biol* 16:2744–2755
- Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, Scribner KT, Bonin A, Fortin MJ (2010a) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Mol Ecol* 19:3760–3772
- Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010b) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Mol Ecol* 19:3824–3835
- Miller NJ, Ciosi M, Sappington TW, Ratcliffe ST, Spencer JL, Guillemaud T (2007) Genome scan of *Diabrotica virgifera virgifera* for genetic variation associated with crop rotation tolerance. *J Appl Entomol* 131:378–385
- Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Mol Ecol* 17:1552–1563
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Mol Ecol* 18:375–402
- Oetjen K, Ferber S, Dankert I, Reusch TBH (2010) New evidence for habitat-specific selection in Wadden Sea *Zostera marina* populations revealed by genome scanning using SNP and microsatellite markers. *Mar Biol* 157:81–89
- Poncet BN, Herrmann D, Gugerli F, Taberlet P, Holderegger R, Gielly L, Rioux D, Thuiller W, Aubert S, Manel S (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, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Reich PB, Wright IJ, Cavender-Bares J, Craine JM, Oleksyn J, Westoby M, Walters MB (2003) The evolution of plant functional variation: traits, spectra, and strategies. *Int J Plant Sci* 164(Suppl.):143–164
- Reusch TBH, Wood TE (2007) Molecular ecology of global change. *Mol Ecol* 16:3973–3992
- Rieseberg LH, Willis JH (2007) Plant speciation. *Science* 317:910–914
- Robertson A (1975) Gene frequency distributions as a test of selective neutrality. *Genetics* 81:775–785

- Roden SE, Dutton PH, Morin PA (2009) AFLP fragment isolation technique as a method to produce random sequences for single nucleotide polymorphism discovery in the Green Turtle, *Chelonia mydas*. *J Hered* 100:390–393
- Schlötterer C (2003) Hitchhiking mapping—functional genomics from the population genetics perspective. *Trends Genet* 19:32–38
- Schmidt PS, Serrão EA, Pearson GA, Riginos C, Rawson PD, Hilbish TJ, Brawley SH, Trussell GC, Carrington E, Wethey DS, Grahame JW, Bonhomme F, Rand DM (2008) Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology* 89(Suppl.):91–107
- Schönswetter P, Stehlik I, Holderegger R, Tribsch A (2005) Molecular evidence for glacial refugia of mountain plants in the European Alps. *Mol Ecol* 14:3547–3555
- Schultze-Motel W (1986) *Arabis alpina*. In: Hegi G (ed) *Illustrierte Flora von Mitteleuropa* V4. Parey, Berlin
- Sgrò CM, Lowe AJ, Hoffmann AA (2011) Building evolutionary resilience for conserving biodiversity under climate change. *Evol Appl* 4:326–337
- Shikano T, Ramadevi J, Merilä J (2010) Identification of local- and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *Mol Biol Evol* 27:2775–2789
- Sobel JM, Chen GF, Watt LR, Schemske DW (2009) The biology of speciation. *Evolution* 64:295–315
- Stinchcombe JR, Hoekstra HE (2007) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100:158–170
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* 14:671–688
- Tedder A, Ansell SW, Lao X, Vogel JC, Mable BK (2011) Sporophytic self-incompatibility genes and mating system variation in *Arabis*. *Ann Bot* 108:699–713
- The Arabidopsis Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Thiel-Egenter C, Alvarez N, Holderegger R, Tribsch A, Englisch T, Wohlgenuth T, Colli L, Gaudeul M, Gielly L, Jogan N, Linder HP, Negrini R, Niklfeld H, Pellecchia M, Rioux D, Schönswetter P, Taberlet P, van Loo M, Winkler M, IntraBioDiv Consortium, Gugerli F (2011) Break zones in the distributions of alleles and species in alpine plants. *J Biogeogr* 38:772–782
- Titz W (1971) Zur evolution der Gattung *Arabis* im Alpenraum. *Ber Deut Bot Ges* 84:697–704
- 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
- Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Mol Ecol* 14:3623–3642
- Verhoeven KJF, Poorter H, Nevo E, Biere A (2008) Habitat-specific natural selection at a flowering-time QTL is a main driver of local adaptation in two wild barley populations. *Mol Ecol* 17:3416–3424
- Wang R, Farrona S, Vincent C, Joecker A, Schoof H, Turck F, Alonso-Blanco C, Coupland G, Albani MC (2009) *PEP1* regulates perennial flowering in *Arabis alpina*. *Nature* 459:423–427
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J Evol Biol* 14:611–619
- Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Mol Ecol* 17:3123–3135
- Wright S (1951) The genetical structure of populations. *Ann Eugen* 15:323–354