# Title: Disentangling adaptation from drift in bottlenecked and reintroduced populations of Alpine ibex

Running title: Disentangling adaptation from drift

Leigh, D.M.,[1]*♦ Lischer, H.E.L.,[2,3] Guillaume, F.,[4] Grossen, C.,[4] Günther, T.[5]
*corresponding author: deborahmleigh.research@gmail.com

**Author information**
1. WSL Swiss Federal Research Institute, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland
2. Interfaculty Bioinformatics Unit, University of Bern, Bern 3001, Switzerland
3. Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland
4. Department of Evolutionary Biology and Environmental Studies, University of Zurich, CH-8057, Zurich, Switzerland
5. Human Evolution, Department of Organismal Biology, Evolutionary Biology Centre, 75236, Uppsala University, Uppsala, Sweden

♦The bulk of this work was carried out while DML and HELL were are the Department of Evolutionary Biology and Environmental Studies, University of Zurich, CH-8057, Zurich, Switzerland

**Orchid IDs:**
Leigh, D.M: 0000-0003-3902-2568
Lischer, H.E.L.: 0000-0002-9616-2092
Guillaume, F.: 0000-0003-0874-0081
Grossen, C.: 0000-0003-4157-1910
Günther, T.: 0000-0001-9460-390X

# Abstract

Identifying local adaptation in bottlenecked species is essential for conservation management. Selection detection methods have an important role in species management plans, assessments of adaptive capacity, and looking for responses to climate change. Yet, the allele frequency changes exploited in selection detection methods are similar to those caused by the strong neutral genetic drift expected during a bottleneck. Consequently, it is often unclear what accuracy selection detection methods have across bottlenecked populations. In this study, simulations were used to explore if signals of selection could be confidently distinguished from genetic drift across 23 bottlenecked and reintroduced populations of Alpine ibex (*Capra ibex*). The meticulously recorded demographic history of the Alpine ibex was used to generate comprehensive simulated SNP data. The simulated SNPs were then used to benchmark the confidence we could place in outliers identified in empirical Alpine ibex SNP data. Within the simulated dataset, the false positive rates were high for all selection detection methods but fell substantially when two or more methods were combined. True positive rates were consistently low and became negligible with increased stringency. Despite finding many outlier loci in the empirical Alpine ibex SNPs, none could be distinguished from genetic drift-driven false positives. Unfortunately, the low true positive rate also prevents the exclusion of recent local adaptation within the Alpine ibex. The baselines and stringent approach outlined here should be applied to other bottlenecked species to ensure the risk of false positive, or negative, signals of selection are accounted for in conservation management plans.



**Keywords:** Outlier, Bottleneck, Reintroduction, Evolutionary management, Conservation

# Introduction

Identification of recent responses to selection, or local adaptation, is of great interest to evolutionary and conservation biologists. Insights gained from recent selective changes can facilitate our understanding of evolutionary processes (Whitlock and Lotterhos, 2015a). For conservation biologists, insights into local adaptation also have a more applied or practical importance. Characterizing within species adaptive differences is often necessary for species management plans (e.g. Robertson *et al.,* 2014), and optimizing source population choice for translocations or reintroductions (Flanagan *et al.,* 2017). Characterizing adaptive processes may also offer insight into long-term extinction risk, particularly if a population or species is no longer able to respond to selection (Frankham *et al.,* 2010). Within reintroduced populations specifically, the sudden environmental change experienced when founder individuals are released in new locations may fuel rapid adaptive change (e.g. Stockwell *et al.,* 2003; Reznick *et al.,* 2004). Understanding of which is important if future- potentially disruptive- translocations are planned. This new conservation ethos where evolutionary processes are considered in species management, is known as "evolutionary" or "adaptive" conservation management (Hoffmann *et al.*, 2015). The long-term success of evolutionary conservation management requires accurate assessments of the evolutionary processes in bottlenecked populations and thus, an understanding of the analytical constraints non-equilibrium populations can face.

The current ease in obtaining genome-wide SNP data has driven a renaissance of studies scanning for selection at the genomic level in wild populations (e.g. *Gasterosteus aculeatus*, Hohenlohe *et al.,* 2010; *Peromyscus maniculatus*, Linnen *et al.,* 2013; *Sarcophilus harrisii,* Epstein *et al.,* 2016; *Oncorhynchus clarkii henshawi,* Amisch *et al.,* 2019). *Fst*-based selection

86 detection methods are widely used to detect recent intra-species selective responses by scanning
87 for unusually high values of *Fst* ("outlier" loci), which are assumed to be driven directly or
88 indirectly (i.e. hitchhiking) by positive selection (Lewontin and Krakauer, 1973; Fay and Wu,
89 2000). Popularity of these methods has fueled analytical extensions that identify selective
90 responses using environmental clines (Coop *et al.,* 2010; De Mita *et al.,* 2013). Referred to as
91 genetic-environment association analyses or "GEA" analyses, these methods pinpoint alleles that
92 display repeated associations with an environmental variable due to local adaptation (Lotterhos
93 and Whitlock, 2015; Hoban *et al.,* 2016). The degree to which currently available selection
94 detection methods successfully accommodate unusual, or more complex demographic histories,
95 is still being tested. This information is essential to ensure accuracy because small demographic
96 assumption violations can fuel elevated rates of false signals of selection, where neutral loci are
97 falsely identified as outliers. This can arise, for example, from unaccounted variance in the
98 distribution of *Fst* due to shared history and relatedness of populations (Robertson, 1975a;
99 Robertson, 1975b; Excoffier *et al.,* 2009). Recent population bottlenecks and reintroductions
100 pose a new challenge for selection detection, because they are associated with very complex
101 patterns of high inter-population relatedness that may violate model assumptions and exacerbate
102 false positive rates (Frankham *et al.,* 2010). Furthermore, the random allele frequency changes
103 caused by the strong genetic drift inherent in a bottleneck can lead to large allele-frequency
104 differences between populations (Kimura 1955a; Kimura 1955b). Genetic drift can therefore
105 create outlier-like loci that can easily be mistaken as loci under selection and will increase the
106 false positive rate of selection detection methods in bottlenecked populations (Lotterhos and
107 Whitlock, 2014; Klopfstein *et al.,* 2006; Nielsen *et al.,* 2007; Foll and Gaggiotti, 2008; Hofer *et*

*al.,* 2009). Such false signals have previously hampered selection scans in bottlenecked species, including humans (Sabeti *et al.,* 2006).

Examination of selection detection accuracy in bottlenecked populations is limited. Foll and Gaggiotti, (2008) examined the effects of including a subset of populations that are bottlenecked in a selection detection analysis. It was recommended to remove bottlenecked populations due to the increase in false positives this caused (Foll and Gaggiotti, 2008). The effects of historical bottlenecks (thousands of generations prior) were also examined using simulated populations of *Peromyscus spp.* (Poh *et al.,* 2014) and *Haemorhous mexicanus* (Shultz *et al.,* 2016), where the false positive rate often exceeded selection detection power. Nevertheless, selection detection analyses have since been applied to bottlenecked populations (e.g. Pilot *et al.,* 2014; Funk *et al.,* 2016; Amish *et al.*, 2019), and will likely continue to be applied, because of the conservation management need to identify intra-species adaptive differences. It is therefore essential that we expand our exploration of bottleneck effects on selection detection accuracy.

The Alpine ibex (*Capra ibex)* is a recently bottlenecked and reintroduced species with a demographic history that is virtually unparalleled in recorded detail (Biebach and Keller, 2009). In this study, we utilized these population records to create a comprehensive simulated SNP data set through individual-based forward simulations. We then examined the performance of different selection detection methods by quantifying both the observed true and false positive rates and the composition of outlier loci. This information was coupled with selection scans on an empirical Alpine ibex restriction site associated DNA sequencing (RADseq) data set, and used to guide the confidence we could place in any outliers detected in these reintroduced populations. This provided insight into the accuracy, or rather lack-there-of, expected within

131 species with complex histories of bottlenecks and reintroductions. The detection thresholds and
132 methods outlined here can be used as a guideline to help avoid false positive loci in other species
133 with similar histories. Furthermore, they highlight the high risk of false negatives, where local
134 adaptation present but unidentified.
135

136 # Materials and methods

137 *Alpine ibex demographic history*

138 Alpine ibex underwent a prolonged decline starting in the 16$^{th}$ century due to overhunting. Only a
139 single population of an estimated 100 individuals survived this crash in the Gran Paradiso region
140 of Northern Italy. Royal protection in the 19$^{th}$ century enabled the population to grow to 3000-
141 5000 individuals. Reintroductions of Alpine ibex from the Gran Paradiso region into Switzerland
142 began in 1906. Detailed demographic records were kept as part of the reintroduction program in
143 Switzerland. Information that was recorded included the origin of founder individuals (often
144 coming from previously reintroduced populations, Figure 1), the number and gender of founders,
145 and the year an individual was moved. In addition, annual census records of the number of
146 animals alive in spring were collected for many reintroduced populations (Stuwe and Grodinsky,
147 1987; Stuwe and Neivergelt, 1991; Biebach and Keller, 2009). This reintroduction program was
148 very successful, to date more than 17 thousand Alpine ibex are present in the Swiss Alps
149 (Shackleton and Group ISCI, 1997; BAFU, 2015; Brambilla *et al.* 2020). The focal populations
150 used in this study are shown in Figure 1.
151

152 *RAD sequencing:*

To apply selection detection methods to an empirical data set from a bottlenecked species, we used the published RADseq data set from Leigh *et al.,* (2018) and Grossen *et al.,* (2017). This consists of 304 Alpine ibex from 23 reintroduced populations (Figure 1). We used only variants called by GATK (Poplin *et al.,* 2017; see Leigh *et al.,* 2018 for a discussion of variant caller effects). After SNP filtering (described in section S3) a sample of 213 individuals remained. For selection detection all singletons were removed and SNPs within 1kb were randomly thinned using vcftools (vcftools; Danecek *et al.,* 2011), which resulted in a final data set of 12695 SNPs. After exclusion of individuals from the Gran Paradiso, inclusion of which potentially violates selection detection analysis relatedness assumptions (Günther and Coop, 2013), 5225 SNPs were suitable for the selection detection analyses.

*Simulating the Alpine ibex history*

Simulated SNP data sets were generated using forward time simulations in Nemo (version 2.3.51; Guillaume and Rougemont, 2006) and used to assess the expected accuracy if each selection detection methods when applied to bottlenecked and reintroduced species. Details of the simulations can be found in the supplementary material (section S1). Briefly, in each simulation all 23 populations sampled for RADseq were simulated. In order to accurately simulate these populations, an additional three populations that were founder sources for the focal populations were also simulated (see panel in Figure 1). Therefore, 26 populations were simulated in total. The reintroduction history and population sizes were informed by detailed records and census data. Ten replicate simulations of the Alpine ibex reintroduction history were conducted for each of three genetic architectures: 1) neutral SNPs only, 2) 30 loci under selection, and 3) 120 loci under selection. The loci under selection were di-allelic QTL

176 contributing additively to a quantitative trait. In all architectures, each individual had 30
177 chromosomes (linkage groups) of 10M (Morgan) each with 60 thousand neutral loci. In the two
178 architectures with selection the 30 or 120 QTL were equally spread among the neutral loci. The
179 recombination rate was $5x10^{-4}$ between adjacent neutral SNPs. The QTL were set either at the
180 center of each chromosome (30 QTL) or four QTL were positioned 3.33M apart and 0.5cM from
181 the start on each chromosome (120QTL). This ensured several thousand SNPs were polymorphic
182 after the bottleneck and generated the same chromosome number and a similar level of linkage
183 disequilibrium to that in the RADseq data set as evaluated by the $r^2$ values between final
184 polymorphic SNPs in vcftools.
185 In each simulation, neutral loci and loci under selection were allowed to reach mutation-
186 selection-drift equilibrium during a "burn-in" of 10 thousand generations in a single population
187 that represented the Gran Paradiso population. After this time, a bottleneck was applied. We
188 simulated phenotypic selection on the quantitative trait with a Gaussian fitness surface where the
189 trait optimum value varies among populations depending on an environmental variable (snow
190 cover). The trait optimum value during the burn-in was held at zero (in the 'Gran Paradiso'
191 reference population) to maintain alleles of both negative and positive effect. To generate post-
192 reintroduction selection across the 30 or 120 QTL, the trait optimum in reintroduced populations
193 was varied to either zero, -2 or +2. Values reflected observed real world snow conditions relative
194 to the Gran Paradiso, for example those with a higher average snow depth had an value of +2 and
195 those with a lower average snow depth had a value of -2. Snow conditions were chosen as they
196 are a strong candidate real-world selection pressure, specifically they have previously been
197 shown to affect Alpine ibex population dynamics and vary dramatically across sites (detailed in

198 Supplementary material S1 and S2 and Table S1 and S2) (Jacobsen *et al.,* 2004; Grøtan *et al.,*
199 2008).

200 The strength of selection at each locus was determined by the size of its contribution to
201 the trait. For the architecture where 30 diploid loci were under selection: six loci had large
202 contributions to each trait (allelic value, $a = \pm 0.1$), and 24 were divided equally into 4 categories
203 of lesser effect ($a = \pm 0.08, \pm 0.04, \pm 0.02, \pm 0.01$). A maximum trait value of ±3 was therefore
204 achievable. For the architecture where 120 loci were under selection, the division of loci
205 remained identical except for the loci of smallest effect. Specifically, 96 loci were of minor
206 effect (±0.01) and 24 were equally divided amongst the remaining allelic values (±0.1, ±0.08,
207 ±0.04, ±0.02, 6 of each value in total). A maximum trait value of ±4.8 was achievable. Selection
208 coefficients ($s$) equaled 0.027 ($a = \pm 0.1$), 0.022 ($a = \pm 0.08$), 0.012 ($a = \pm 0.04$), 0.007 ($a = \pm 0.02$)
209 and 0.004 ($a = \pm 0.01$) in both architectures. This was calculated according to Bürger (2000)
210 using the phenotypic variance ($V_s$) of 0.047 (120 loci under selection) or 0.035 (30 loci under
211 selection), as well as a selection variance ($\omega^2$) of 7.5. This generated two biologically realistic
212 trait architectures and realistic strengths of selection.

213 The simulated genotypes from the final generation were used to evaluate the expected
214 accuracy of different selection detection methods, and only polymorphic SNPs were included in
215 the simulated data from this time point. To mimic the available RADseq data, 10 simulated
216 individuals were randomly chosen from each of the 23 populations that were sequenced with
217 RADseq, 6000 polymorphic loci were the taken for each individual including all polymorphic
218 selected loci and a subset of randomly selected neutral loci. 20% of genotypes were randomly set
219 to "missing" due to missing data in the RADseq genotypes and singletons were removed

220 (vcftools; Danecek *et al.,* 2011). PGDspider (version: 2.0.9.2; Lischer and Excoffier, 2012) and
221 custom scripts were used to convert Nemo output into input for the selection analyses.
222
223 *Screens for signals of positive selection*
224 Selection detection analyses were conducted for both the empirical Alpine ibex RADseq data
225 and simulated data sets. This enables us to quantify the confidence we could place in any
226 empirical outliers. To detect signatures of selection, Bayenv 2.0 (Günther and Coop, 2013),
227 Baypass 2.1 (Gautier, 2015a), and OutFLANK (Whitlock and Lotterhos, 2015a) were used
228 (following Leigh *et al.,* 2018). These three programs were chosen as they have been shown to
229 have high accuracy in species with complex patterns of population relatedness (Günther and
230 Coop, 2013; Lotterhos and Whitlock, 2014; Gautier, 2015a; Whitlock and Lotterhos, 2015a).
231 Bayenv 2.0 and Baypass2.1 utilize a modified *Fst*-like statistic called $X^TX$ that is corrected for
232 shared population history (Günther and Coop, 2013; Gautier, 2015a). Outflank utilizes an *Fst*
233 statistic called *F'st*, a metric based on Wright's *Fst* statistic without corrections for a finite
234 sample size (Whitlock and Lotterhos, 2015a). These three methods are hereafter referred to as
235 *Fst*-like approaches. Bayenv 2.0 and Baypass2.1 also detect selection using GEA selection scans
236 (as in Hoban *et al.,* 2016).
237 Selection detection program conditions are detailed in Leigh *et al.,* (2018). Briefly, the
238 estimation of covariance matrix and subsequence selection scan in Bayenv 2.0 were run
239 independently three times with 2 $x10^5$ Markov-Chain-Monte-Carlo (MCMC) iterations (Blair *et*
240 *al*., 2014). SNPs were considered putatively under selection for the GEA method, if the Bayes
241 factor (BF) value exceeded 3 and the Spearman's rho value was in the top and bottom 2.5% of
242 all SNPs across the three runs. This threshold was chosen because it suggests high support for a

SNP being under selection and that the trend is not due to a single outlier population (Nadeau *et al.,* 2016; Günther and Coop, 2013). The *Fst*-like approach SNPs had to have $X^TX$ value among the top 100 ranking SNPs across all three runs (Günther and Coop, 2013).

Baypass2.1 was run three times for each data set with 20 pilot runs of 1000 MCMC iterations and 5000 MCMC iterations for the "burn-in" (default conditions). For the GEA analysis we used the Auxillary model and consider a loci to be under selection when it had a 10 x $\log_{10}$ Bayes factor (db) greater than 4.7 for all three replicates (Gautier, 2015a). This value is equivalent to the threshold of a BF of 3 used in Bayenv 2.0. For the *Fst*-like approach, $X^TX$ outliers were determined following the best-practice tutorial accompanying Baypass2.1 (Gautier, 2015b). This uses trained-simulations to find the 99% threshold for $X^TX$ values for each dataset, outliers were those loci in the top 1% for all three Baypass runs (Gautier, 2015b).

In OutFLANK, outlier SNPs were identified following the best practice tutorial (default settings, Whitlock and Lotterhos, 2015b). To be considered an outlier, a SNP had to have a Q-value of less than 0.05 (Storey and Tibshirani, 2003; Whitlock and Lotterhos, 2015a), as well as a heterozygosity of greater than 10% (Whitlock and Lotterhos, 2015b).

Loci identified across multiple programs as outliers were also compared. Loci identified as outliers across two programs were called "double positives" those found by all three programs were called "triple positives." To account for the different signals the *Fst*-like and GEA approaches look for, the outliers identified by the two methods in Bayenv 2.0 and Baypass2.1 were not combined into a single set. Thus we had double and triple positive *Fst*-like outliers, and double positive GEA outliers. For the triple positive GEA outliers, the GEA outliers from Bayenv 2.0 and Baypass2.1 were overlapped with the *Fst*-like outliers from OutFLANK because OutFLANK does not us a GEA approach.

All environmental data used in the GEA analyses were obtained from MeteoSwiss (Switzerland). For each population, data from the closest meteorological station available (Figure 1, Section S1 and S2, Table S2) were used to obtain averages since a population was founded, or since records began. The environmental variables in the analyses were divided across winter and summer and included air temperature (°C), daily precipitation (mm), and snow depth measures (cm). Further details are available in the supplementary material (section S1 and S2). Since the simulations were intended to mimic real Alpine ibex populations, the corresponding weather data were included as environmental covariates in the Bayenv 2.0 and Baypass2.1 analyses of the simulated data. In addition, each simulated population's true simulated environmental optimum was also included as an environmental covariate in the analysis of the simulated data (Table S1).

*Evaluating method accuracy with simulations*

The simulated genotype data were used to estimate the true or false negative and positive rates. When examining loci flagged as putatively under selection, a true positive was considered to be a simulated locus under selection that was correctly identified as being under selection. A false positive was considered to be a simulated neutral locus that was wrongly identified as being under selection. The proportion of all loci identified by a method as under selection that were true positives, hence indeed under selection (the true discovery rate), was used as a metric of the method's accuracy and reliability of selection detection. To place the results in the context of other simulation studies, the true positive rate, false positive rate, the false discovery rate, and false negative rate, were also calculated. All metrics are defined in Table 1 for ease of reference. All values displayed are the averages across 10 simulated datasets for each genetic architecture

and are relative only to the number of polymorphic QTL loci and neutral loci in the final SNP set.

# Results

In this study, we generated empirical RADseq and simulated SNP data for the Alpine ibex. Bayenv 2.0, Baypass2.1, and OutFLANK were then used to identify loci putatively under selection in these datasets. The simulated data provided an estimate of the selection detection accuracy of these three popular tools in the empirical Alpine ibex dataset. Low true discovery rates were identified for all selection detection methods (detailed below), preventing us from confidently distinguishing selection from false positive outliers in the Alpine ibex RADseq data.

*Alpine ibex RADseq data and signals of selection:*

Each selection detection method identified outliers in the Alpine ibex RADseq data set. Between 172 to 2 loci were found to be putatively under selection by the different selection detection methods (Figure 2A). However, only 14 loci were identified as double positives and no locus exceeded the triple positive threshold. The highest number of double positive loci was found by the Bayenv Baypass GEA overlap. The two other double positive loci were found separately in the overlap of Bayenv and Baypass *Fst*-like, as well as the Bayenv and Outflank *Fst*-like overlap. As detailed below, this is within the range of drift-driven false positives expected under all simulated genetic architectures.

*Evaluating expected selection detection accuracy*

Analyses of simulated data revealed a very low selection detection accuracy under the Alpine ibex demography, regardless of the genetic architecture simulated. Figure 2B shows the false positive rates for the neutral only simulations and Figure 3 the true and false discovery rates (i.e. the composition of loci identified as outliers) for the simulations with loci under selection. For the two architectures with selection, the true positive rate, false positive rate and false negative rates are shown in Table 2.

For all simulation types, each individual selection detection method had a high number of false positives and a striking false negative rate (Figures 3). The false positives rate did decrease considerably (<0.001) for the double and triple positive methods, but this was at the expense of the false negative rate increasing (Table 2). Greater variability in accuracy is seen for the architecture with 30 loci under selection than 120 loci under selection. Specifically, the true discovery rate does occasionally reach 1.0 (see Figure 3). However, as shown by the true positive rate and false negative rate (Table 2), this does not reflect high accuracy of these methods but stochastic chance. Virtually all simulations had no outliers exceed this threshold, but a single simulation had 1 true positive locus, leading to a mean true discovery rate of 1.

In the simulations with selection, the allelic values and hence the strength of selection experienced by each QTL locus, were not equal. The loci with allelic values of 0.1 or 0.08 were under much stronger selection ($s$=0.027, 0.022) relative to those with allelic values of 0.04, 0.02 or 0.01 ($s$=0.012, 0.007, 0.004). Consequently, the signal of selection and therefore the true positive rate may be unequal across loci under selection. Table 3 shows the average allele frequency change of loci under selection, this can be considered a rough proxy for the signal of selection visible at a locus. As expected due to the strength of selection, loci under the strongest

selection were often at extreme allele frequencies after the burn-in and before the bottleneck (Figure S1 and S2). Consequently, such loci were fixed more frequently over the course of our simulations and thus more likely to be excluded from selection detection analysis. Nevertheless, loci under a selection pressure of >0.022 were the most likely to be identified as outliers in the architecture with 30 loci under selection. Those under weaker selection (0.004) were most likely to be identified as outliers in the architecture with 120 loci under selection, but this was because they were by far the most common in this architecture and their abundance drives this trend.

# Discussion

In this study the accuracy of selection detection methods was assessed for the Alpine ibex, a species with a complex history of bottlenecks and reintroductions. We generated comprehensive simulations that followed the species' recorded population history. Three genetic architectures were simulated: neutral loci only, 30 loci under selection, and 120 loci under selection. The simulated data revealed a low selection detection accuracy for each individual selection detection method. Improved accuracy was possible when only considering outliers identified by multiple methods, though this came at the expense of an increased false negative rate. This made it impossible to adjust our thresholds as we were either overrun with false positives, or rarely identified ongoing selection. While candidate outlier loci could be identified in the Alpine ibex RADseq data set, the simulation results indicate they cannot be confidently considered as under selection. Importantly, the low true positive rate also prevents us from confidently concluding the absence of recent adaptation in the populations, posing significant challenges for the evolutionary management of this species. Nevertheless, identifying false positive outliers and

355 concluding two populations are separate Evolutionary Significant Units has a number of costly
356 consequences for conservation management. Until more accurate selection detection methods are
357 found, the stringent approach and criteria outlined here should be applied to other bottlenecked
358 species to offer an indication of the confidence that we can place in outlier loci.
359 *Screen for selection with Alpine ibex RADseq data*
360 In the Alpine ibex RADseq dataset 14 loci were identified as under selection using the double
361 positive approach but no loci were triple positives. Based on the simulations, a proportion of
362 <0.04 of loci identified by the double positive approach are likely to be true positives. This
363 extremely low proportion indicates that these putatively selected loci should be viewed with
364 extreme caution because many are likely to be false positive loci. Consequently, these loci were
365 not explored further (as in, Shultz *et al.,* 2016). Interestingly, the significant environmental
366 correlations found in the GEA outliers were related to environmental variables known to have
367 recruitment effects and to vary dramatically across the reintroduced range. Despite biologically
368 realistic explanations, the expected high rates of false positives prevent us from making any
369 confident conclusions about local adaptation in the Alpine ibex at this time. Furthermore, the size
370 and nature of this species make the functional validation that was used in *Peromyscus spp.*
371 impossible (Poh *et al.,* 2014). Though it is likely some adaptation may be occurring in Alpine
372 ibex, these candidate outliers and those found in other bottlenecked species, must be confirmed
373 when more accurate selection methods for bottleneck population are identified in the future.
374 Future studies should focus on selection detection methods less reliant on *Fst* (e.g. time series
375 approach, Brüniche-Olsen *et al*., 2016), and explore if sufficient power can be gained by more
376 densely sampling the genome with Whole Genome Sequencing (Lowry *et al.,* 2017). For studies
377 interested in examining multiple naturally bottlenecked populations (i.e. not reintroduced

species) exploiting museum and collection specimens could also be used to circumvent major genetic drift driven false positives by offering pre-bottleneck allele frequencies.

*Simulated data and selection detection accuracy:*

Alpine ibex have experienced several profound and serial population bottlenecks. Given this extreme history, genome-wide drift effects are highly likely and a high false positive rate was expected for selection detection methods applied to this data (Kimura 1955a; Kimura 1955b; Lotterhos and Whitlock, 2014). The simulations of the Alpine ibex demography confirmed this, revealing an expected false positive rate of up to 0.03 and a false discovery rate often exceeding 0.99 of all outliers. This accuracy was considerably less than that found for non-bottlenecked populations and for scans where a single population is bottlenecked (e.g. 0.1 false positive rate, Foll and Gaggiotti, 2008). However, the low accuracy is similar to studies where more ancient bottlenecks were simulated (e.g. 0.03-0.41 false positive rate, Poh *et al.,* 2014; 0.05-0.30, Shultz *et al.,* 2016). Importantly, increasing stringency to a double or triple positive approach did improve the false positive rate in the Alpine ibex data. This suggests that the double or triple overlap approaches may offer some improved power in bottlenecked populations, and their accuracy should be assessed for more simple bottleneck scenarios. However, this approach increases the already high risk of being too stringent and removing all loci under selection (high false negative rate), which must also be taken in to account when applying this method.

A low true positive rate was identified for all simulated loci under selection. To generate a biologically realistic trait, majority of loci simulated were of small or moderate effect and it has been previously demonstrated that many selection detection methods struggle to identify such loci, regardless of demographic history (e.g. Biswas and Akey, 2006; Kalsson and Moen, 2010;

401 Narum and Hess, 2011; Kemper *et al*., 2014; Lotterhos and Whitlock, 2015). This is particularly
402 pronounced for loci contributing to polygenic traits such as ours (Kemper *et al.*, 2014; Berg and
403 Coop, 2014). However, in this study, loci under comparable selection coefficients were identified
404 much less frequently than expected based on previous studies. Specifically in our study, loci with
405 a selection coefficient below 0.012 were rarely identified by the double or triple positive method.
406 However, Lotterhos and Whitlock (2015) found a true positive rate of at least 0.11 for loci under
407 a weaker selection coefficient of 0.005, with two or more selection detection methods. Our true
408 positive rate for loci of the largest effect was also lower than seen previously, for example for the
409 Bayenv GEA we found a 0.04 true positive rate, while previous studies have found 0.58-1 across
410 multiple demographic scenarios (Coop *et al.,* 2010; De Mita *et al.,* 2013; Lotterhos and
411 Whitlock; 2015).
412 The lower accuracy found here is likely driven by a combination of factors, including the
413 intrinsic characteristics of bottlenecked populations. Specifically, the swamping of true positives
414 with drift-driven false positives (which will increase the false discovery and false positive rate),
415 as well as the lower effective population size of a bottlenecked species. A lower effective
416 population size will reduce the efficacy of selection (Frankham *et al.,* 2010). This in turn limits
417 detectable signals of selection. Though 17 thousand Alpine ibex are now present in the Alps,
418 population connectivity is low and contemporary population sizes are often in the hundreds.
419 Effective population sizes range from ~900 to as low as 20 (Biebach and Keller, 2009). While
420 the strength of selection at loci with an allelic value of 0.1 or 0.8 ($s$>0.02) was sufficient to
421 theoretically elicit a response even in the smallest simulated populations ($s$>1/2Ne, Frankham *et*
422 *al.,* 2010), loci of the smallest effect will not overpower drift unless the effective population size
423 exceeds 125 individuals and the census size of three of our simulated populations fell below this

threshold. The reduced efficacy of selection in our smallest populations must disrupt signals of selection at loci under weak selection, and contribute to the low true positive rate observed for these loci. In addition, loci under stronger selection were more often at extreme allele frequencies after the burn-in (i.e. preceding any bottleneck) and their rare alleles were easily lost during the bottlenecks or during the shifts in selection pressures. Many of these loci had to be subsequently excluded from selection scans due to their fixation across all populations, exacerbating our difficulty in identifying selection. These issues are likely common to selection scans on bottlenecked species where selection is long acting (i.e. continuous before and during a bottleneck). Accordingly, true positive rate is similar to that found in other bottlenecked species (e.g. Poh *et al.,* 2014). This is highly problematic for adaptive population management, because long standing adaptive differences are often exactly what we are trying to conserve. This result does suggest that greater success may be had when looking for signals of post-bottleneck adaptation, for example when scanning for rapid post-reintroduction adaptation to a novel environmental variable or adaptation to a new disease. To circumvent the reduced accuracy due to fixation of selected alleles, future studies should explore if any increase in power is obtained through using pre-bottleneck samples for SNP ascertainment.

*Conclusions*

Overall, for populations like the Alpine ibex with a history of extreme population bottlenecks (and notably, serial founding events as well as complex reintroductions) the selection detection methods explored here have a considerably reduced accuracy relative to other demographic histories. Based on these results, loci identified as under selection in similar bottlenecked populations using GEA or *Fst* outlier methods should be viewed with caution, particularly those

based on single selection detection methods. Unfortunately for bottlenecked species, the high false positive rate is also coupled with a high false negative rate. Therefore, if selective responses are not identified in bottlenecked populations, this cannot be considered evidence for an absence of responses to selection pressures or an absence of local adaptation. This unfortunate lack of power is highly problematic for effective adaptive population management and it is vital this uncertainty is now incorporated into management plans. Alongside this, the costs of concluding two populations as separate ESUs based on erroneous outliers must be evaluated. The criteria and approach outlined here, may offer other studies on bottlenecked species an approach and baseline on which to gauge their confidence in any outliers identified and adjust management plans accordingly. In the future, the accuracy of selection detection methods less reliant on *Fst,* such as those exploiting temporal samples, as well as use of more dense marker data, should be evaluate across bottlenecked scenarios. Despite the high false positive rate expected, it is important to see if these approaches offer greater power and if they can better facilitate conservation management.

# Acknowledgements

This project was funded by a University of Zurich's Research Priority Program "Evolution in Action" grant to Lukas F Keller and Andreas Wagner. We would like thank them both for their support throughout this project. We would also like to thank the European Science Foundation for funding a Short Visit Grant to DML and TG to establish the collaboration. We would like to thank the ETH's Genetic Diversity Center and BSSE for their help generating the RADseq data.

We would also like to thank Glauco Camenisch, Kasia Sluzek and Iris Biebach for their help during the project.

# References

Amish, S.J., Ali, O., Peacock, M., Miller, M., Robinson, M., Smith, S., …, Neville., H., (2019) Assessing Thermal Adaptation Using Family−based Association and F ST Outlier Tests in a Threatened Trout Species. Molecular Ecology 28(10): 2573–93. https://doi.org/10.1111/mec.15100.

BAFU (2015) Abschuss Steinbock, ganze Schweiz: 2000-2015. Retrieved from: www.wild.uzh.ch/jagdst/index.php

Berg, J. J., & Coop, G. (2014). A Population Genetic Signal of Polygenic Adaptation. PLoS Genetics, 10(8).1-25. https://doi.org/10.1371/journal.pgen.1004412https://doi.org/10.1371/journal.pgen.1004412

Biebach, I., & Keller, L.F. (2009). A strong genetic footprint of the re-introduction history of Alpine ibex (*Capra ibex ibex*). Molecular Ecology. 18, 5046–5058. 10.1111/j.1365-294X.2009.04420.x

Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. Trends in Genetics, 22(8), 437–446. https://doi.org/10.1016/j.tig.2006.06.005

Brambilla, A., Hardenberg, A.V., Nelli, L., & Bassano, B. 'Distribution, Status, and Recent Population Dynamics of Alpine Ibex Capra Ibex in Europe'. Mammal Review 50, no. 3 (2020): 267–77. https://doi.org/10.1111/mam.12194.

Brüniche-Olsen, A., Austin, J. J., Jones, M. E., Holland, B. R., & Burridge, C. P. (2016). Detecting selection on temporal and spatial scales: A genomic time-series assessment of selective responses to devil facial tumor disease. PLoS ONE, 11(3), 1–15. doi.org/10.1371/journal.pone.0147875

Coop, G., Witonsky, D., Rienzo, A. D., & Pritchard, J. K. (2010). Using Environmental Correlations to Identify Loci Underlying Local Adaptation. Genetics, 185(4), 1411-1423 doi.org/10.1534/genetics.110.114819 doi.org/10.1534/genetics.110.114819

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.,A., …Wang, J. (2011). The variant call format and VCFtools. Bioinformatics, 27,2156–8. 10.1093/bioinformatics/btr330

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. Molecular Ecology, 22(5), 1383–1399. https://doi.org/10.1111/mec.12182https://doi.org/10.1111/mec.12182

Epstein, B., Jones, M., Hamede, R., Hendricks, S., McCallum, H., Murchison, E. P., … Storfer, A. (2016). Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nature Communications, 7, 12684. https://doi.org/10.1038/ncomms12684https://doi.org/10.1038/ncomms12684

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. Heredity, 103(4), 285–98. https://doi.org/10.1038/hdy.2009.74

Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. Genetics, 155(3), 1405–1413.

Flanagan, S. P., Forester, B.R., Latch, E.K., Aitken, S.N., Hoban, S. (2018) Guidelines for Planning Genomic Assessment and Monitoring of Locally Adaptive Variation to Inform Species Conservation. Evolutionary Applications 11(7): 1035–52. https://doi.org/10.1111/eva.12569.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics, 180(2), 977–993. https://doi.org/10.1534/genetics.108.092221https://doi.org/10.1534/genetics.108.092221

Frankham, R., Ballou, J., & Brisco, D. (2010). Introduction to Conservation Genetics (2nd ed.). UK: Cambridge University Press

Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., … Andelt, W. F. (2016). Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). Molecular Ecology, 25(10), 2176–2194. https://doi.org/10.1111/mec.13605

Gautier, M. (2015a). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. Genetics 201,1555–79. 10.1534/genetics.115.181453

Gautier, M. (2015b). Baypass version 2.1 User Manual. Retrieved from: http://www1.montpellier.inra.fr/CBGP/software/baypass/index.html

Grossen, C., Biebach, I., Angelone-Alasaad, S., Keller, L. F., & Croll, D. (2017). Population genomics analyses of European ibex species show lower diversity and higher inbreeding in reintroduced populations. Evolutionary Applications.11:123–139

Grøtan, V., Saether, B.E., Filli, F., & Engen, S. (2007). Effects of climate on population fluctuations of ibex. Global Change Biology, 14, 218–228 10.1111/j.1365-2486.2007.01484

Guillaume, F., & Rougemont, J. (2006). Nemo: an evolutionary and population genetics programming framework. Bioinformatics, 22, 2556–2557 10.1093/bioinformatics/btl415

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. Genetics, 195(1), 205–220. https://doi.org/10.1534/genetics.113.152462

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. The American Naturalist, 188(4), 379–397. https://doi.org/10.1086/688018

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. Annals of Human Genetics, 73(1), 95–108. https://doi.org/10.1111/j.1469-1809.2008.00489.x

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics, 6(2). https://doi.org/10.1371/journal.pgen.1000862

Jacobson, A., & Provenzale, A. (2004). Climate forcing and density dependence in a mountain ungulate population. Ecology. 85,1598–1610. doi/10.1890/02-0753

Karlsson, S., & Moen, T. (2010). The power to detect artificial selection acting on single loci in recently domesticated species. BMC Research Notes, 3, 232. https://doi.org/10.1186/1756-0500-3-232

Kemper, K. E., Saxton, S. J., Bolormaa, S., Hayes, B. J., & Goddard, M. E. (2014). Selection for complex traits leaves little or no classic signatures of selection. BMC Genomics, 15(1), 246. https://doi.org/10.1186/1471-2164-15-246

Kimura, M. (1955a). Solution of a Process of Random Genetic Drift With a Continuous Model. Proceedings of the National Academy of Sciences of the United States of America, 41(3), 144–50. https://doi.org/10.1073/pnas.41.3.144

Kimura, M. (1955b). Random Genetic Drift in Multi-Allelic Locus. Evolution, 9(4), 419–435. 10.2307/2405476

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. Molecular Biology and Evolution, 23(3), 482–490. https://doi.org/10.1093/molbev/msj057https://doi.org/10.1093/molbev/msj057

Leigh, D. M., Lischer, H. E. L, Grossen, C., & Keller, L. F. 'Batch Effects in a Multiyear Sequencing Study: False Biological Trends Due to Changes in Read Lengths'. Molecular Ecology Resources 18, no. 4 (2018): 778–88. https://doi.org/10.1111/1755-0998.12779.

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics, 74(1), 175–195.

Linnen, C. R., Poh, Y.P., Peterson, B. K., Barrett, R. D. H., Larson, J. G., Jensen, J. D., & Hoekstra, H. E. (2013). Adaptive Evolution of Multiple Traits Through Multiple Mutations at a Single Gene. Science, 339(6125), 1312–1316. https://doi.org/10.1126/science.1233213https://doi.org/10.1126/science.1233213

Lischer, H.E.L., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 28, 298–299. 10.1093/bioinformatics/btr642

Lotterhos, K.E., & Whitlock, M.C. (2014). Evaluation of demographic history and neutral parameterization on the performance of Fst outlier tests. Molecular Ecology. 23, 2178–2192. 10.1111/mec.12725

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Molecular Ecology, 24(5), 1031–46. https://doi.org/10.1111/mec.13100

Lowry, D. B., Hoban, S., Kelley, J.L., Lotterhos, K.E., Reed, L.K., Antolin, M.F. & Storfer, A. Breaking RAD: An Evaluation of the Utility of Restriction Site-Associated DNA Sequencing for Genome Scans of Adaptation. Molecular Ecology Resources 17, no. 2 (March 2017): 142–52. https://doi.org/10.1111/1755-0998.12635.

Nadeau, S., Meirmans, P. G., Aitken, S. N., Ritland, K., & Isabel, N. (2016). The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. Ecology and Evolution, 6(24), 8649–8664. https://doi.org/10.1002/ece3.2550https://doi.org/10.1002/ece3.2550

Narum, S. R., & Hess, J. E. (2011). Comparison of F(ST) outlier tests for SNP loci under selection. Molecular Ecology Resources, 11 Suppl 1, 184–94. https://doi.org/10.1111/j.1755-0998.2011.02987.x

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., & Clark, A. G. (2007). Recent and ongoing selection in the human genome. Nature Reviews Genetics, 8(11), 857–868. https://doi.org/10.1038/nrg2187

Pilot, M., Greco, C., vonHoldt, B. M., Jędrzejewska, B., Randi, E., Jędrzejewski, W., … Wayne, R. K. (2014). Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. Heredity, 112(4), 428–442. https://doi.org/10.1038/hdy.2013.122

Poh, Y.P., Domingues, V.S, Hoekstra, H. E. & Jensen, J. D. 'On the Prospect of Identifying Adaptive Loci in Recently Bottlenecked Populations'. PLOS ONE 9, no. 11 (10 November 2014): e110579. https://doi.org/10.1371/journal.pone.0110579

Poplin, R., Ruano-Rubio, V., DePristo,M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., … Banks, E. (2017) Scaling accurate genetic variant discovery to tens of thousands of samples *bioRxiv*

Reznick, D., Rodd, H., & Nunney, L. (2004) Empirical evidence for rapid evolution. In: Ferrière, R., Dieckmann, U., Couvet, D. ed. Evolutionary Conservation Biology. UK: Cambridge University Press

Robertson, A. (1975a). Gene frequency distributions as a test of selective neutrality. Genetics, 81(4), 775–785.

Robertson, A. (1975b). Letters to the editors: Remarks on the Lewontin-Krakauer test. Genetics, 80(2), 396.

Robertson, J.M. Langin, K.M, Scott Sillett T., Morrison S.A.,Ghalambor, C.K., Funk, W.C. (2014) Identifying Evolutionarily Significant Units and Prioritizing Populations for Management on Islands. Monographs of the Western North American Naturalist, 7(1), 397-411

Sabeti, P. C. (2006). Positive Natural Selection in the Human Lineage. Science, 312(5780), 1614–1620. https://doi.org/10.1126/science.1124309

Stockwell, C.A., Hendry, A.P., & Kinnison, M.T. (2003). Contemporary evolution meets conservation biology. Trends in Ecology Evolution. 18, 94–101 dx.doi.org/10.1016/S0169-5347(02)00044dx.doi.org/10.1016/S0169-5347(02)00044-7

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genome-wide studies. Proceedings of the National Academy of Sciences, 100(16), 9440–9445. https://doi.org/10.1073/pnas.1530509100

Shackleton, D., & Group ISCS. (1997). Wild Sheep and Goats and their Relatives. Status Survey and Conservation Action Plan for *Caprinae.* IUCN, Gland, Switzerland and Cambridge.

Shultz, Allison J., Allan J. Baker, Geoffrey E. Hill, Paul M. Nolan, and Scott V. Edwards. 'SNP s across Time and Space: Population Genomic Signatures of Founder Events and Epizootics in the House Finch (*Haemorhous Mexicanus* )'. Ecology and Evolution 6, no. 20 (October 2016): 7475–89. https://doi.org/10.1002/ece3.2444.

Stuwe, M., & Grodinsky, C. (1987). Reproductive biology of captive Alpine ibex (*Capra i. ibex*). Zoo Biology. 6, 331–339. 10.1002/zoo.1430060407

Stuwe, M., & Nievergelt, B. (1991). Recovery of Alpine ibex from near extinction – the result of effective protection, captive breeding, and reintroductions. Applied Animal Behaviour Science. 29,379–387.

Whitlock, M. C., & Lotterhos, K. E. (2015a). Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of *FST.* The American Naturalist, 186(S1), S24–S36. https://doi.org/10.1086/682949https://doi.org/10.1086/682949

Whitlock, M. C., & Lotterhos, K. E. (2015b). OutFLANK Finding *Fst* outliers with an inferred neutral distribution. https://github.com/whitlock/OutFLANK

# Data accessibility

Read data can be viewed on the short-read archive, ncbi project number PRJNA422727: https://www.ncbi.nlm. nih.gov/bioproject/PRJNA422727.

# Author contributions

DML performed the selection detection analysis, simulations and wrote the manuscript

TG supported the selection detection analysis and commented on the manuscript.

CG supported the sequence data generation, commented on the manuscript and simulations.

FG supported writing the simulation scripts and designing the genetic architecture of the QTL traits.

# Figures and Tables

***Figure 1:*** *The 23 Alpine ibex focal populations and a simplified representation of the reintroduction history in Switzerland equating to the effective bottleneck number each population experienced (top left panel). All Swiss populations descend from the Gran Paradiso national park in Northern Italy (open circle), which is included in the figure but was excluded from the selection detection analysis. Reintroductions in Switzerland often used founder individuals from previously established reintroduced populations. As a result, many populations have experienced several serial bottlenecks. Within this figure, each circle represents a Swiss Alpine ibex focal population and the circle's shading indicates the number of bottlenecks each population experienced. Marked by a cross are the weather stations used to estimate the local environment experienced by each population.*

***Figure 2: A)*** *The number of empirical outliers detected by each selection detection method in the Alpine ibex RADseq SNP set.* ***B)*** *The false positive rate from the fully neutral simulations. Shown below each bar is the average number of outlier loci identified*

***Figure 3:*** *The true and false discovery rate of different selection detection methods for* ***A)*** *the architecture with 30 loci under selection and* ***B)*** *the architecture with 120 loci under selection. Each bar shows the average composition of loci identified as outliers using each selection detection method, at the bottom of the bar is the average number of outliers across 10 replicate simulations. Replicates where no loci exceeded the significance threshold were excluded from the figure.*

***Table 1:*** *Definitions of each metric used to assess a selection detection method's accuracy with the simulated data.*

| Accuracy metric | Definition |
|---|---|
| *True discovery rate* | The proportion of all simulated loci identified as outliers that were actually under selection (i.e. QTL loci). |
| *True positive rate* | The proportion of loci under selection (i.e. QTL loci) correctly identified as an outlier. |
| *False positive rate* | The number of neutral loci incorrectly identified as under selection (false positive outliers) divided by the number of retained polymorphic neutral SNPs. |
| *False discovery rate* | The proportion of outlier SNPs that were false positives (i.e. simulated neutral loci)(Lotterhos and Whitlock, 2014) |
| *False negative rate* | The proportion of polymorphic QTLs that were not identified as outliers (and thus not identified as under selection). |

***Table 2:*** *Selection detection accuracy as measured by the true and false positive rate, as well as the false negative rate. 30 or 120 signifies the number of loci under selection* (*QTL loci).*

| Selection detection method | True positive rate | | False positive rate | | False negative rate | |
|---|---|---|---|---|---|---|
| | **30** | **120** | **30** | **120** | **30** | **120** |
| Bayenv *Fst-like* | 0.026 | 0.012 | 0.012 | 0.012 | 0.974 | 0.988 |
| Baypass *Fst-like* | 0.005 | 0.001 | 0.004 | 0.005 | 0.995 | 0.999 |
| OutFLANK *Fst-like* | 0.009 | 0.001 | 0.000 | 0.001 | 0.991 | 0.999 |
| Bayenv GEA | 0.005 | 0.004 | 0.003 | 0.003 | 0.995 | 0.996 |
| Baypass GEA | 0.033 | 0.030 | 0.032 | 0.031 | 0.967 | 0.970 |
| Double positive Bayenv Baypass *Fst-like* | 0.005 | 0.001 | 0.001 | 0.001 | 0.995 | 0.999 |
| Double positive Bayenv OutFLANK *Fst-like* | 0.009 | 0.001 | 0.000 | 0.001 | 0.991 | 0.999 |
| Double positive Baypass OutFLANK *Fst-like* | 0.005 | 0.001 | 0.000 | 0.001 | 0.995 | 0.999 |
| Double positive Bayenv Baypass GEA | 0.000 | 0.003 | 0.001 | 0.001 | 1.000 | 0.997 |
| Double positive Bayenv OutFLANK GEA | 0.005 | 0.000 | 0.000 | 0.000 | 0.995 | 1.000 |
| Double positive Baypass OutFLANK GEA | 0.005 | 0.000 | 0.000 | 0.000 | 0.995 | 1.000 |
| Triple positive *Fst-like* | 0.005 | 0.001 | 0.000 | 0.001 | 0.995 | 0.999 |
| Triple positive GEA | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 |

675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690

***Table 3:*** *Mean absolute allele frequency change for loci under selection ± the standard error. Shown in brackets is the percentage of loci that remain polymorphic in at least one population at the end of the simulations. Values are calculated from immediately after the burn-in using the values from the simulated Gran Paradiso population, relative to the frequency across all simulated populations in final generation. Loci fixed after the burn-in were excluded from the values.*

692
693

| | **Average allele frequency change (percentage polymorphic)** | |
|---|---|---|
| **Locus type** | **30 Loci under selection** | **120 Loci under selection** |
| 0.01 | 0.086±0.071 (93%) | 0.087±0.079 (86%) |
| 0.02 | 0.095±0.082 (94%) | 0.093±0.078 (93%) |
| 0.04 | 0.067±0.067 (77%) | 0.088±0.078 (83%) |
| 0.08 | 0.043±0.058 (48%) | 0.045±0.060 (49%) |
| 0.1 | 0.031±0.034 (44%) | 0.016±0.022 (23%) |

694
695
696
697
698
699
700
701
702
703

# Supplementary material

**Section S1:** Simulations of neutral loci and loci under selection

**Section S2:** Weather Data

**Section S3:** RADseq bioinformatics steps

**Table S1:** The simulated optima of the populations included in the simulations. Positive and negative values were based on the relative difference in snow depth between each population and the Gran Paradiso population. An optimum of zero was used for the burn-in

**Table S2:** Weather stations used for each population. Data included in the analyses are an average of each season over the years since each population's founding or when records began. Stations were chosen based on proximity and similarity in conditions.

**Figure S1**: Allele frequencies of selection loci from the architecture with 30 loci under selection. Shown are loci that remain polymorphic after the burn-in within the remnant population (left panels) and their subsequent frequencies across all populations (right panels) in the final simulated generation. Top panels are loci under strong selection with an allelic value of 0.08 or 0.1, bottom panels are loci under weaker selection with an allelic value of 0.04,0.02 or 0.01.

**Figure S2:** Allele frequencies of selection loci from the architecture with 120 loci under selection. Shown are loci that remain polymorphic after the burn-in within the remnant population (left panels) and their subsequent frequencies across all populations (right panels) in the final simulated generation. Top panels are loci under strong selection with an allelic value of 0.08 or 0.1, bottom panels are loci under weaker selection with an allelic value of 0.04,0.02 or 0.01.