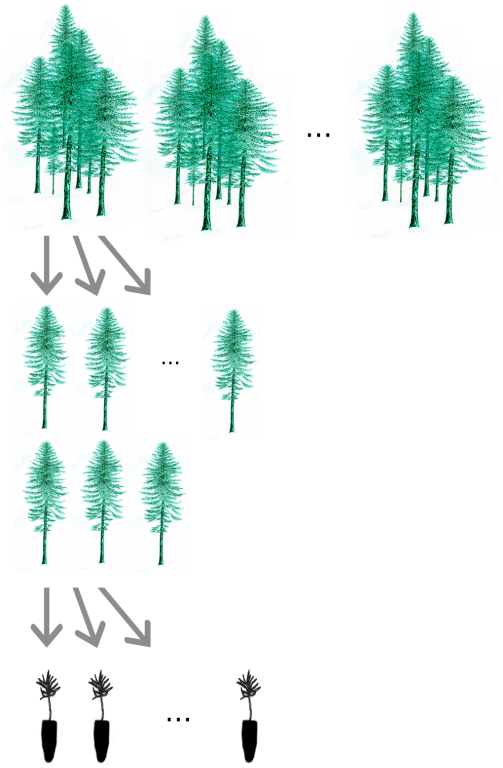
Supplementary Information

Title: Adaptation to local climate in a multi-trait space: evidence from silver fir (*Abies alba* Mill.) populations across a heterogeneous environment

Authors: Katalin Csilléry, Otso Ovaskainen, Christoph Sperisen, Nina Buchmann, Alex Widmer, Felix Gugerli

The following supplementary information are available for this article:



Experimental units and samples taken	Data obtained	Statistical analysis
Region Selection of 19 putatively autochthonous silver fir populations	Coarse-scale data on pastclimate Fine scale data for recent climate	Downscaling coarse scale (~55 km) past climate data to a 1 km grid scale
Population Needle samples from 20 adult trees	SNP genotypes $\delta^{13}C$	Population genetic differentiation (F_{ST}) Demographic inference: 1. coancestry matrix (AFM) 2. genetic clustering (STRUCTURE) Mean water use efficiency of populations
Mother tree Seeds from 3 mother trees (some included in the 20)	Weight of 1000 seeds (maternal effects)	Establishing the pedigree
Half-sib family ~14 seedlings grown in a common garden	Quantitative traits in the 4^{th} and 5^{th} growing season	Population differentiation in quantitative traits (Q_{ST})

Combined analysis of selection using multiple data types:

1. Q_{ST}-F_{ST} comparisons
2. Bayesian animal model to estimate additive genetic traits values and population effects while simultaneously accounting for the demography and the pedigree
3. S-test of selection in quantitative traits
3. H-test for the role of climate in trait divergence
4. Contrasting seedling trait divergence and adult mean water use efficiency

Figure S1: Schematic illustration of the work-flow.

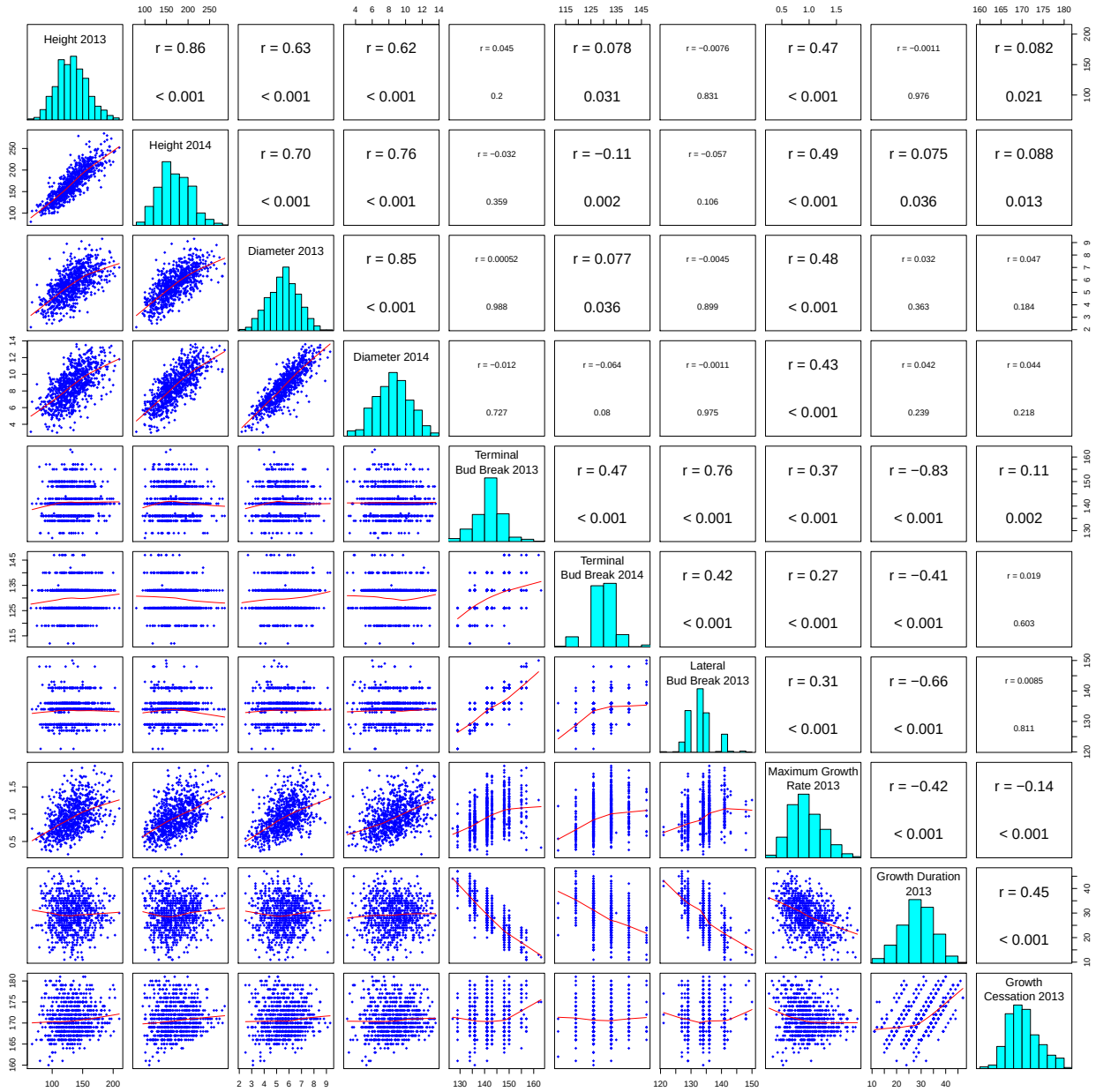


Figure S2: Raw values of traits measured in a common garden on silver fir (*Abies alba* Mill.) seedlings. Figures show scatter plots between values for all pairs of traits (lower triangle), the distribution of each trait (diagonal), and the Pearson's correlation coefficient and associated *p*-value from a correlation test between values of all pairs of traits (upper triangle).

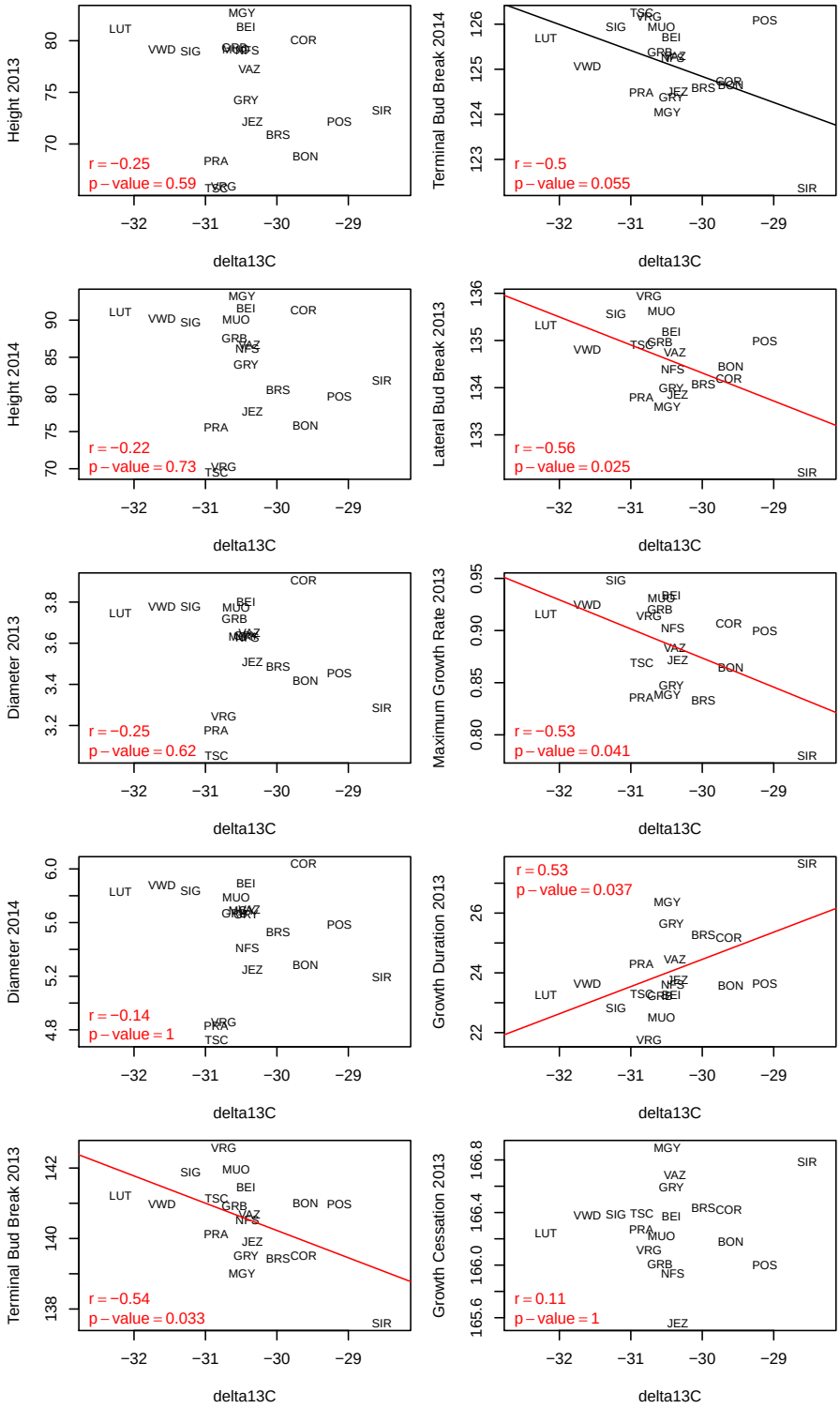


Figure S3: Correlation between population mean $\delta^{13}C$ of adult trees measured *in-situ* and the additive genetic trait values of seedlings in the ten quantitative traits. Pearson's correlation and p-values from a correlation test adjusted for multiple testing are written on each panel in red. The direction of significant correlations is indicated with a red line, while a black line indicates a marginally significant correlation.

Table S1: Abbreviations of the population names, their longitude and latitude, elevation in meters, and names of the nearby village (all in Switzerland) from which the population name abbreviations were derived. Data base/Publication/Expert specifies the source of information for deriving the putatively autochthonous status of the populations (see next page for details of abbreviations). NKS status is provided for populations present in the NKS data base.

Population code	Latitude	Longitude	Elevation	Nearby village	Data base/Publication/Expert	NKS status
BEI	47.230°*N*	8.318°*E*	843	Beinwil	Dr A. Burkhart (WSL, Researcher)/NKS	la
BON	46.324°*N*	9.541°*E*	1334	Bondo	Huss/Burga	NA
BRS	46.595°*N*	6.175°*E*	1221	Le Chenit (Le Brassus)	Huss/Burga	NA
COR	47.162°*N*	7.055°*E*	840	Cormoret	NKS	la
GRB	47.334°*N*	9.114°*E*	922	Oberhelfenschwil (Graben)	IUFRO/NKS	a
GRY	46.299°*N*	7.091°*E*	1433	Gryon	NKS	la
JEZ	46.921°*N*	9.700°*E*	1158	Jenaz	U. Bḧler (GR, Forest Office)	NA
					M. Flury (Jenaz GR, District forester)	
LUT	46.634°*N*	7.952°*E*	817	Lütschental	Burga/K. Zumbrunn (BE, Forester)	NA
MGY	46.095°*N*	7.100°*E*	1022	Martigny	C. Pernstich (VS, Forestry Office)	NA
MUO	46.991°*N*	8.708°*E*	691	Muotatal	Max Bc̈hel (Muotatal SZ, District forester)	NA
NFS	47.090°*N*	8.997°*E*	1152	Näfels	J. Walcher and	NA
					K. Winzeler (GL, Forestry Office)	
POS	46.270°*N*	10.082°*E*	1602	Poschiavo (Le Prese)	U. Bühler (GR, Forestry Office)	NA
PRA	46.479°*N*	8.750°*E*	1180	Prato (Leventina)	Huss/Burga/IUFRO	NA
SIG	46.891°*N*	7.761°*E*	938	Signau	Huss/Burga/Dr L. Walthert (WSL, Researcher)	NA
SIR	46.280°*N*	7.560°*E*	1149	Sierre	IUFRO/NKS/Huss	a
TSC	46.938°*N*	10.481°*E*	1284	Tschlin	IUFRO/NKS/Huss/Burga	a
VAZ	46.639°*N*	7.002°*E*	965	Maules	Dr L. Walthert (WSL, Researcher)	NA
VRG	46.237°*N*	8.530°*E*	1149	Vergeletto	IUFRO/NKS/Huss	la
VWD	47.273°*N*	7.884°*E*	481	Vordemwald	LWF/Huss/	NA
					Dr L. Walthert and Dr P. Weber (WSL, Researcher)	NA

NKS: Der nationale Kataster der Samenerntebestnde (National catalogue of seed sources) http://www.nks.admin.ch/
IUFRO: International Union of Forest Research Organizations
https://www.iufro.org/
LWF: Long-term Forest Ecosystem Research
https://www.wsl.ch/en/forest/forest-development-and-monitoring/long-term-forest-ecosystem-research-lwf.html
Huss: Hussendörfer, E. (1997): Untersuchungen über die genetische Variation der Weisstanne (*Abies alba* Mill.) unter dem Aspekt der In-situ-Erhaltung genetischer Ressourcen in der Schweiz (PhD thesis ETH Zurich Nr. 11849). Beih. Schweiz. Z. Forstwes. 83: 1–151 (in German)
Burga: Burga, CA and Hussendörfer, E (2001): Vegetation history of *Abies alba* Mill. (silver fir) in Switzerland–pollen analytical and genetic surveys related to aspects of vegetation history of *Picea abies* (L.) H. Karsten (Norway spruce). Vegetation History and Archaeobotany 10 (3): 151-15
WSL: Swiss Federal Research Institute WSL
a: autochthonous
la: likely autochthonous
NA: not applicable

Table S2: Principal Component Analysis (PCA) of the 36 environmental variables listed in Table 1 (main text). PC axes 1 to 5 explained 84.5% of the variance in the raw environmental variables. The first ten environmental variables with the highest loadings are shown for each PC axes.

PC1 (38.1%)		PC2 (18.0%)		PC3 (14.0%)		PC4 (9.5%)		PC5 (5.0%)	
Variables	Loadings	Variables	Loadings	Variables	Loadings	Variables	Loadings	Variables	Loadings
late.frost2	-0.26	bio.16	-0.36	Long	0.35	bio.4	-0.38	bio.15	0.56
PET.thorn	-0.26	bio.12	-0.35	bio.14	-0.33	bio.7	-0.37	late.frost	-0.35
bio.1	-0.26	bio.13	-0.35	SPEI.q5	-0.32	scPDSI.m3	0.26	bio.8	0.32
bio.6	-0.26	bio.18	-0.33	bio.8	0.24	bio.15	0.26	scPDSI.q1	0.31
Elevation	0.26	PET.harg	0.29	scPDSI.m3	0.24	SPEI.m1	0.25	bio.17	-0.29
bio.10	-0.25	SPEI.m2	-0.28	scPDSI.q5	-0.24	bio.9	-0.21	scPDSI.m4	-0.25
bio.11	-0.25	SPEI.q1	0.25	Lat	0.23	scPDSI.q5	-0.21	bio.19	-0.23
bio.3	0.25	bio.17	-0.21	bio.19	-0.21	bio.19	0.2	bio.18	0.15
bio.5	-0.25	SPEI.m1	0.18	SPEI.m1	-0.2	SPEI.m2	-0.2	bio.13	0.15
AWC	0.23	bio.7	0.17	scPDSI.q1	-0.2	bio.2	-0.19	SPEI.m1	0.13

Table S3: *S*-test, genotypic (r_g) correlations estimated from the posterior distribution of the ancestral *G*-matrix (see Results in the main text for details), phenotypic correlations (r_p), the absolute difference between the latter two (Abs. diff.), and the p-value from a standardized Mantel test that indicates if the G_A and *P*-matrices are correlated with each other. Traits were measured on silver fir (*Abies alba* Mill.) seedlings in a common garden.

Trait pair	*S*-test	r_g			r_p	Mantel-test	
		Mean	Lower 95% CI	Upper 95% CI	Mean	Abs. diff.	p-value
Height 2013-Diameter 2013	1.00	0.31	0.23	0.38	0.63	−0.32	0.000
Height 2013-Terminal Bud Break 2013	1.00	0.04	−0.09	0.15	0.04	0.00	0.023
Height 2013-Lateral Bud Break 2013	1.00	−0.01	−0.13	0.12	−0.01	0.00	0.009
Height 2013-Maximum Growth Rate 2013	1.00	0.28	0.13	0.39	0.47	−0.19	0.000
Height 2013-Growth Duration 2013	1.00	0.01	−0.08	0.11	0.00	0.01	0.026
Height 2013-Growth Cessation 2013	1.00	0.06	−0.06	0.22	0.08	−0.02	0.001
Diameter 2013-Terminal Bud Break 2013	0.98	0.00	−0.16	0.16	0.00	0.00	0.000
Diameter 2013-Lateral Bud Break 2013	0.99	0.03	−0.14	0.21	0.00	0.02	0.001
Diameter 2013-Maximum Growth Rate 2013	0.88	0.37	0.19	0.52	0.48	−0.11	0.003
Diameter 2013-Growth Duration 2013	0.98	0.00	−0.15	0.12	0.03	−0.03	0.001
Diameter 2013-Growth Cessation 2013	0.87	−0.04	−0.25	0.18	0.05	−0.01	0.059
Terminal Bud Break 2013-Lateral Bud Break 2013	0.94	1.00	1.00	1.55	0.76	0.55	0.557
Terminal Bud Break 2013-Maximum Growth Rate 2013	0.89	0.51	0.16	0.84	0.37	0.13	0.000
Terminal Bud Break 2013-Growth Duration 2013	0.92	−0.93	−1.09	−0.68	−0.83	0.10	0.000
Terminal Bud Break 2013-Growth Cessation 2013	0.90	0.07	−0.25	0.38	0.11	−0.04	0.104
Lateral Bud Break 2013-Maximum Growth Rate 2013	0.91	0.56	0.25	0.86	0.30	0.25	0.002
Lateral Bud Break 2013-Growth Duration 2013	0.95	−0.91	−1.10	−0.66	−0.66	0.25	0.063
Lateral Bud Break 2013-Growth Cessation 2013	0.93	−0.05	−0.38	0.30	0.01	0.04	0.516
Maximum Growth Rate 2013-Growth Duration 2013	0.88	−0.42	−0.61	−0.20	−0.42	0.00	0.000
Maximum Growth Rate 2013-Growth Cessation 2013	0.65	−0.15	−0.48	0.24	−0.14	0.00	0.001
Growth Duration 2013-Growth Cessation 2013	0.92	0.29	0.03	0.53	0.45	−0.16	0.037
Height 2014-Diameter 2014	1.00	0.31	0.23	0.37	0.76	−0.45	0.000
Height 2014-Terminal Bud Break 2014	1.00	−0.05	−0.13	0.04	−0.11	−0.06	0.010
Diameter 2014-Terminal Bud Break 2014	0.93	−0.04	−0.20	0.14	−0.06	−0.02	0.000

Supplementary Methods S1: Estimating of the heritability, evolutionary potential and population differentiation across the 19 populations: the effect of experimental design, covariates, sample size and scaling

Effect of nursery design on quantitative genetic parameters

The full common garden study of (Frank *et al.*, 2017) consisted of 4107 observations on 91 populations and 259 families. The subset analyzed in this paper, for which genetic marker and water use efficiency data had been collected consisted of 880 observations on 19 populations and 57 families. Seedlings were planted in seven nursery beds each with 47 row pairs. Provenances were planted in the order of reading rows of an imaginary grid placed on the map of Switzerland from the top left to right bottom corner. The three families were planted one after the other, each taking up two rows. Size, growth and phenology traits were not measured in the nursery. Height and Diameter were first measured in 2012 after the transplantation to the common garden site, thus these measures may also incorporate differing planting depth. In contrast, the weight of 1000 seeds were measured for each family, which provides means of accounting for maternal effects.

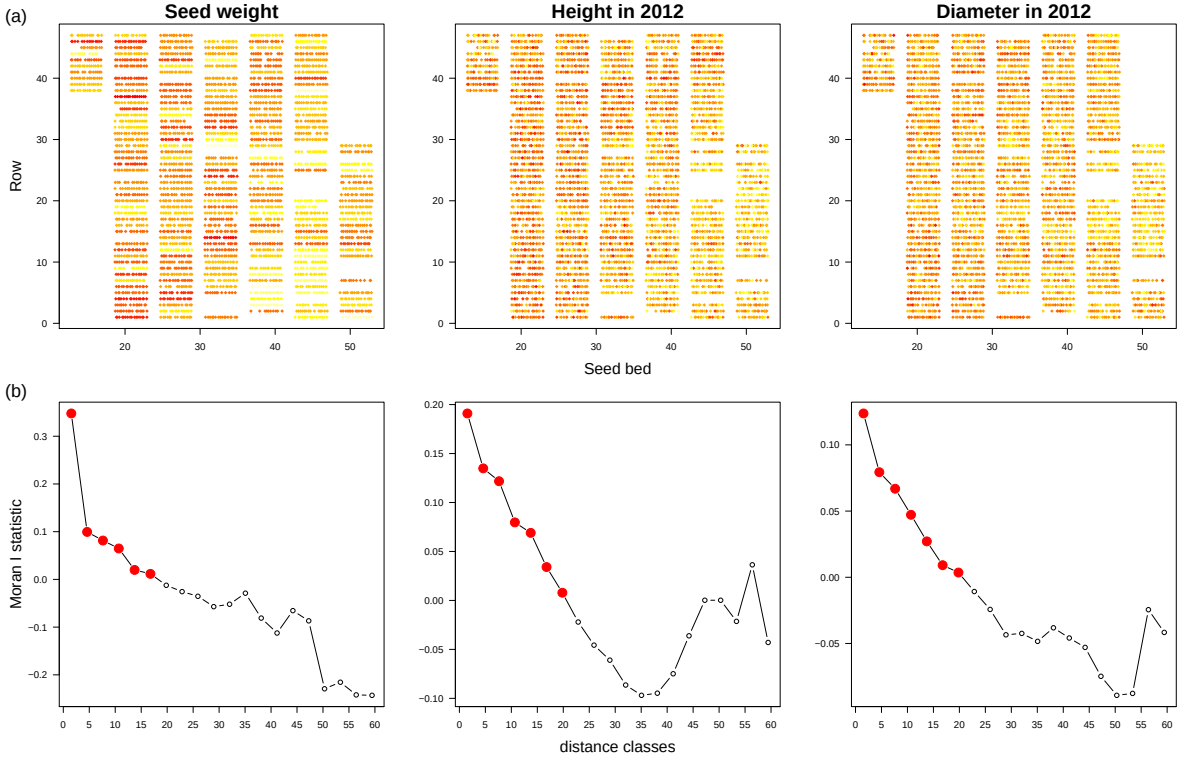


Figure A1: (a) Spatial patterns in Seed Weight, Height 2012 and Diameter 2012 in the nursery. x (seed bed) and y (row) coordinates correspond to distances in meters. The 94 rows are illustrated by 47 row-pairs that the three families of a population occupied. Colors reflect the absolute values of each variable with larger values being darker red. (b) Spatial autocorrelation with respect to the spatial arrangement in the nursery expressed with Moran's I. Red dots indicate significant clustering of similar values.

Fig. A1(a) shows maps of the nursery beds with colors proportional to Seed Weight (trait names are capitalized hereafter), and seedling's Height and Diameter 2012 and the degree of

spatial autocorrelation (Fig. A1(b)). Although there is clear evidence for spatial autocorrelation in Height and Diameter 2012 related to the spatial arrangement in the nursery (Fig. A1), we argue that this is entirely due to population and family effects that were non-randomized. The two lines of evidence to support this are: (i) the presence of spatial autocorrelation with respect to spatial arrangement in the nursery is present already for Seed Weight, and (ii) the lack of spatial correlation other than the effects of population and family for Height 2012 (Table A1). Thus, we conclude that it is likely that the non-randomization of populations and families did not alter the consecutive trait measures. Yet, including Seed Weight as a covariate in models of 2013 and 2014 traits seems useful because it potentially accounts for maternal effects.

Frank *et al.* (2017) used Height 2012 as a covariate to account for nursery effects when estimating population and family variance components for all traits measured in 2013 and 2014. Height 2012 is strongly correlated with Height 2013 (Pearson's correlation of r=0.85), thus including it as a covariate, does not account for the temporal correlation due to repeated measures for Height. Thus, including Height 2012 as a covariate decreased the population and family variance components for Height and Diameter 2013 and Growth variables, while leaving the phenology traits unaffected (Fig. A2(a)). Thus, we did not follow the statistical treatment used in Frank *et al.* (2017).

Effect of sample size on quantitative genetic parameters

First, we assessed the effect of reduced sample size, i.e. using a subset of 19 populations out of the 91. We were interested if quantitative genetic parameters were sensitive to number of populations used, in other words, if they represent similar amounts of trait variation than the larger sample. Population genetic parameters estimated using the reduced set were overall similar to those obtained using the larger data set (Fig. A2(b)). CV_A was the most robust sample size, while Q_{ST} estimates were generally larger, which was expected given that we selected a subset of populations representing diverse ecological conditions. The ordering of the traits in terms of h^2 and Q_{ST} were also different, which most likely reflects the choice of particular populations.

Second, we assessed the effect of having only three families per population, which is the bear minimum for estimating population genetic differentiation. The low number of families also affected the study of Frank *et al.* (2017), however, they partly compensated for it by having a large number of populations (so-called "genecological" approach). We used the full data set of 91 populations to "borrow" families from nearby populations, thereby increase the number of families. We select populations that are nearby our 19 populations taking into account the following criteria: populations had to be no further than 35 km from each other, they had to be in same valley or in the neighbouring valley with the same exposition, and had to have no

Table A1: Model comparison of mixed effects models fitted with lme in R with family nested in population as random effects and with and without Seed Weight as a covariate and with and without spatial autocorrelation structure (SpAC).

Seed Weight	SpAC	df	AIC	BIC	logLik	Ratio	p-value
no	no	4	32840	32865	−16416		
no	yes	6	32843	32881	−16416	1	0.700
yes	no	5	32796	32827	−16393		
yes	yes	7	32799	32843	−16393	1	0.677

more than 200 m difference in elevation. According to these criteria, we were able to increase the number of families in 9 populations, so that the average number of families was 5.3 (range: 3-12). Our analysis showed that the resulting parameter estimates were extremely similar for all three parameters to those obtained with three families only (Fig. A2(c)).

Is the additive genetic variance homogeneous across populations?

Estimating the additive genetic variance from populations across a heterogeneous landscape involves the assumption that the additive genetic variance is constant across the sampling area. Indeed, the global population parameters presented on Fig. A2(a), and thus the results presented in Frank *et al.* (2017), assume a common additive genetic variance across Switzerland. In order to test if such an assumption is reasonable, we estimated two standardized measures of the additive genetic variance, h^2 and CV_A, for the main geographic regions of Switzerland separately using the full data set of 91 populations.

Switzerland is divided into six main forestry regions based on a phylogeographic study (Burga & Hussendörfer, 2001). We pooled these regions to three regions to assure that each region have a sample size of at least 880 to make it comparable with our study using 19 populations. As a result, the three regions had 1247 (Plateau), 1699 (Alps), and 884 (South) observations. We found that h^2 and CV_A were relatively consistent across regions (Fig. A3). Not surprisingly, CV_A was more consistent across regions than h^2 because the latter is dependent on the environmental variance, which is likely different among the regions. These results suggest that regardless of the different demographic history and potential lack of gene flow between the regions separated by high mountain passes, the additive genetic variance is of similar magnitude across the landscape across the different traits.

In conclusion, assessment of the effect of geographic region and sample size highlights that assuming a common additive genetic variance across either 90 or 19 populations without accounting for their demography and potentially differing selection pressures stays relatively robust, but not free from potential biases. Overall CV_A was more stable with respect to sample size, family number and region. Thus, in agreement with Houle (1992), Hansen *et al.* (2011), the mean standardized measure of the additive genetic variance seems more appropriate as a measure of the evolutionary potential and for comparative purposes.

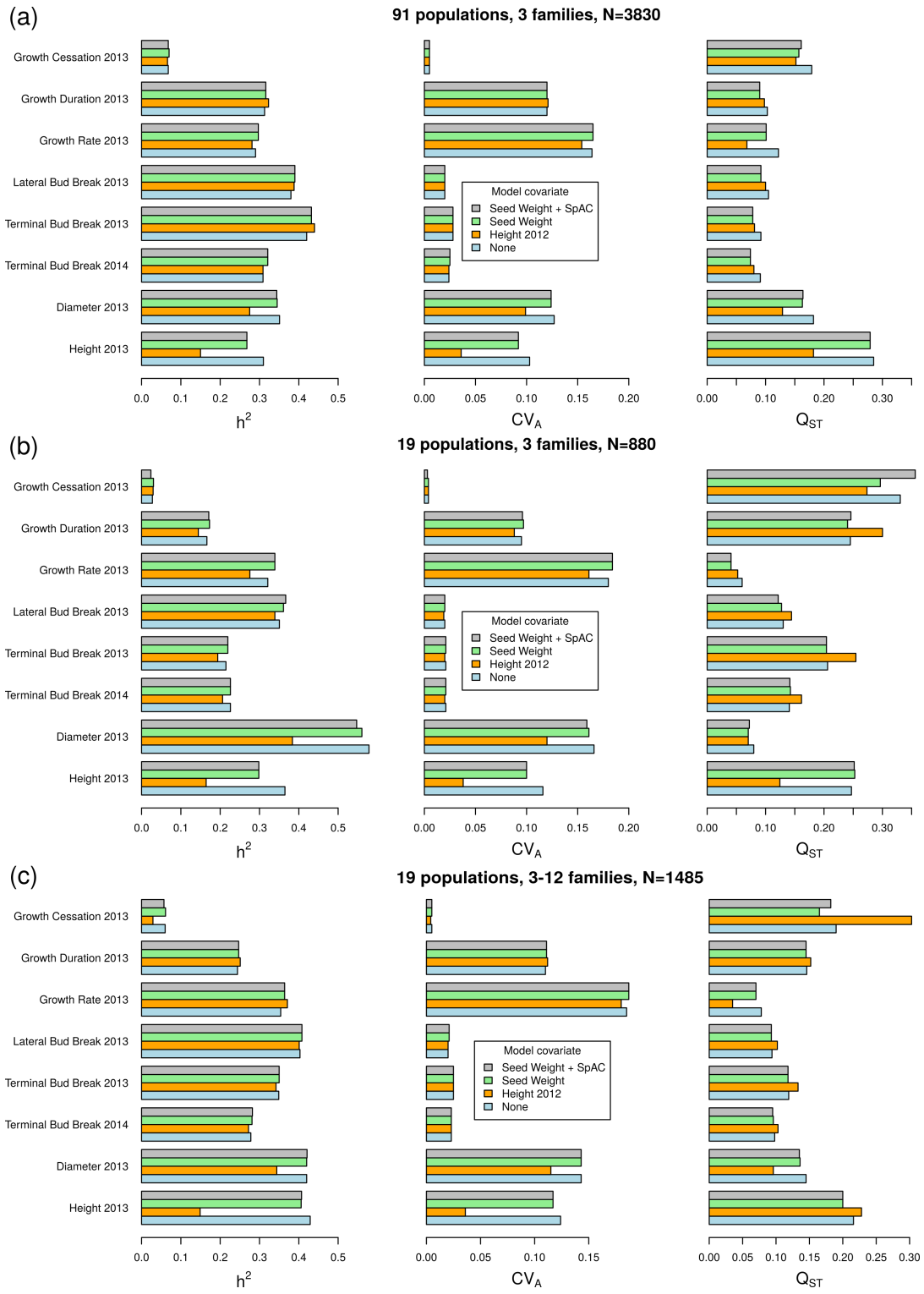


Figure A2: Heritability (h^2), additive genetic coefficient of variation (CV_A) and genetic differentiation between populations (Q_{ST}) estimated using four different versions of a mixed effects model fitted with lme in R with family nested in population as random effects and block in the common garden as fixed effect, and covariates as indicated in the legend. SpAC stands for spatial autocorrelation structure. The model with covariate Height 2012 was used in Frank *et al.* (2017). (a) using the full data set from Frank *et al.* (2017); (b) using the 19 populations studied herein; (c) using the 19 populations studied herein, but with increasing the number of families per population by pooling nearby populations.

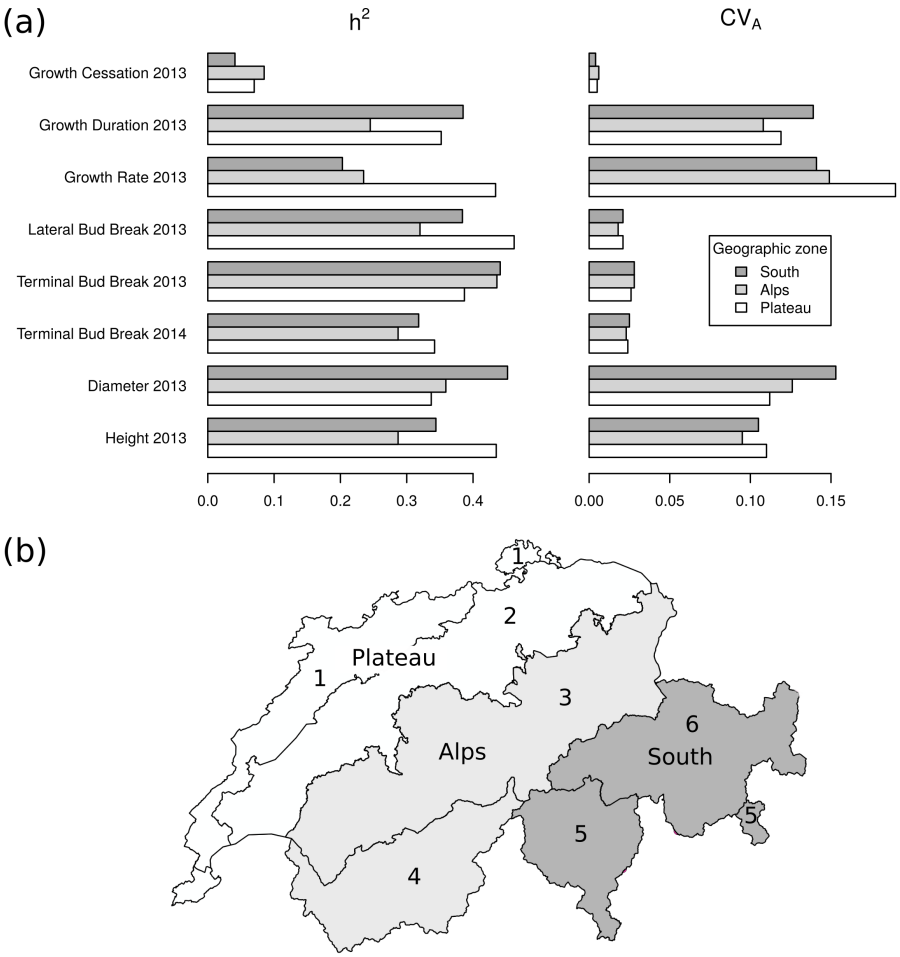


Figure A3: (a) h^2 and CV_A estimated using only the populations from one of the three main geographic regions of Switzerland. (a) The six forestry regions of Switzerland are indicated by numbers (1: Jura, 2: Plateau, 3: Alps, 4: Valley, 5: Ticino, 6: Grison). We pooled these six regions to three to have at least 800 observations for each region. The names of three regions are labelled and marked with different colours. The matching colour coding of (a) and (b) helps the reader.

Supplementary Methods S2: Estimating the demography of the 19 silver fir populations

Estimating the demography of 19 populations is a high dimensional estimation problem. We used the admixture F model (AFM), which assumes that the current populations are derived from a single, non-sampled, ancestral population (Karhunen & Ovaskainen, 2012). The method of (Ovaskainen *et al.*, 2011) also relies on this assumption. For parameter estimation, a Metropolis-Hastings algorithm is implemented in the R package RAFM (Karhunen & Ovaskainen, 2012). We ran ten independent Markov chains of the AFM model using a burn-in of 30,000 iterations followed by 10,000 iterations for estimation with a thinning interval of ten. The estimated posterior distribution of the coancestry matrix is 19 by 19, and it is challenging to use directly this matrix for convergence diagnostics. Thus, we calculated the F_{ST} from each matrix to calculate the convergence diagnostics using the R package *coda* (Fig. A4). Single chain diagnostics indicated satisfactory convergence (Table A2). All chains passed Heidelberger's test (Heidelberger & Welch, 1981). Geweke's statistics were calculated using the default window sizes, 0.1 and 0.5. We found that z-scores were between -1.96 and 1.96, indicating convergence (Geweke, 1991).

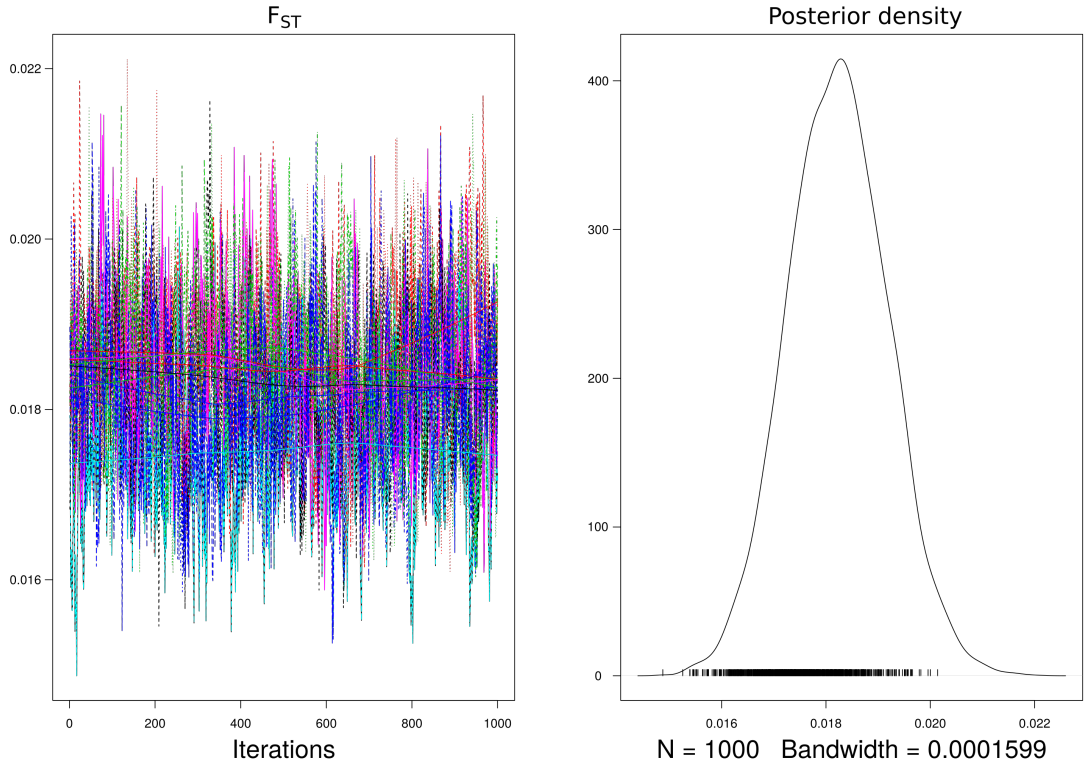


Figure A4: The posterior distribution of F_{ST} calculated from the posterior distribution of the 19 by 19 coancestry matrix across ten independent chains of the Metropolis-Hastings algorithm.

Mixing of the chains was assessed using Gelman's mean potential scale reduction factor (Gelman & Rubin, 1992). A value of one indicate that the variance between and within chains is equal. Different Markov chains reached slightly different optima, in particular, two chains had a lower F_{ST} than the others (Fig. A4 and A5). The mean potential scale factor was 1.09, with an upper credible interval of 1.19. Thus, we estimated the posterior mean coancestry matrix from each Markov chain and averaged them across the ten Markov chains.

Table A2: Single chain convergence diagnostics. The posterior distribution of F_{ST} was calculated from the posterior distribution of the 19 by 19 coancestry matrix across ten independent chains of the Metropolis-Hastings algorithm.

Chain	Heidelberger and Welch		Geweke
	result	p-value	z-score
1	passed	0.63	−1.21
2	passed	0.30	0.25
3	passed	0.74	0.53
4	passed	0.78	1.07
5	passed	0.63	−1.21
6	passed	0.05	1.81
7	passed	0.22	1.31
8	passed	0.78	−0.37
9	passed	0.50	−0.67
10	passed	0.09	0.86

To further validate the results of the AFM model, we compared it to the Bayesian clustering algorithm implemented in STRUCTURE v.2.3.4 Pritchard *et al.* (2000). We used the admixture model with correlated allele frequencies Falush *et al.* (2003), which is the closest model to AFM. Further, we included sampling location information to improve clustering performance ("locprior model", Hubisz *et al.* (2009)), an additional information that cannot be accounted for in AFM. We estimated the prior population allele frequency parameter (λ) from the data, as the default of 1 is not necessarily a good choice for SNP data, where most minor alleles are rare. We estimated λ using K = 1 to avoid non-identifiability issues with the other hyper-parameters (λ, α, F). λ was consistently around 0.65 across ten repeated runs (range: 0.63-0.66, median: 0.65). Then, we tested K values from 1 to 19 using ten independent Markov chains for each K, and 500,000 burn-in iterations and 500,000 iterations for estimation of the membership coefficients. Different number of clusters (K) were compared with Structure Harvester (Earl *et al.*, 2012) using the $LnPr(X|K)$ and Evanno *et al.*'s (2005) method. Admixture coefficients were averaged across ten repeated runs using CLUMPP v.1.1.2 Jakobsson & Rosenberg (2007) using the Greedy algorithm for any K>3 and large-K-Greedy for K>5.

STRUCTURE and AFM results were compared for each K using a Mantel test using the *ecodist* R package (Goslee & Urban, 2007). We calculated a distance matrix from the coancestry matrix and compared to a distance matrix calculated from the CLUMPP outfiles for each K. The individual coancestries of the CLUMPP outfiles were first reduced to population coancestries by taking the mean coancestry of the individuals within a population for each K yielding a 19 $\times$ K matrix for $K = 2, \therefore, 19$. Dendrograms from AFM and STRUCTURE were also compared visually using the R package *dendextend* (Galili, 2015).

The software STRUCTURE using the log likelihood criteria suggested five as optimal number of clusters, nevertheless, additional increase in the log likelihood is suggestive of deeper hierarchical structure (Figure A6). The highest similarity between AFM and STRUCTURE was achieved for $K = 4$ (Figure A7 and A8). STRUCTURE also generally confirmed the East-West differentiation as first level structure and several parts of the dendrograms from the two algorithms were identical (Figure A8). For comparison, the mean Mantel statistic between all pairs of coancestry matrices from the ten independent chains was 0.928 (range: 0.854 - 1). Thus, on average, the ten different AFM chains were more

similar to each other than AFM to STRUCTURE, there were AFM chains just as similar to each other as AFM to STRUCTURE.

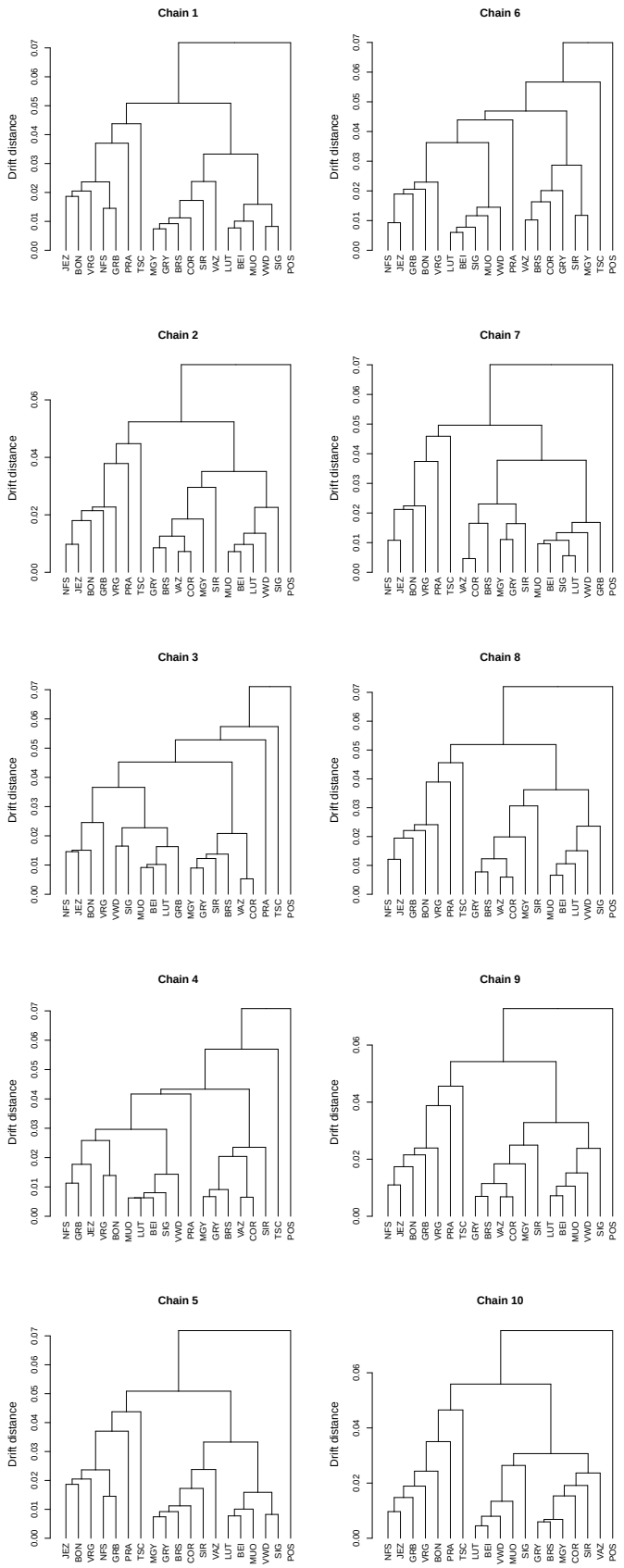


Figure A5: Drift distances between populations estimated using ten independent Markov chains of the AFM model. Each panel corresponds to a Markov chain. Distances were calculated from the posterior mean coancestry matrix to draw the dendrogram. Note that the final coancestry between populations used for inference of adaptive divergence in the main text was the mean of the posterior means from ten independent Markov chains.

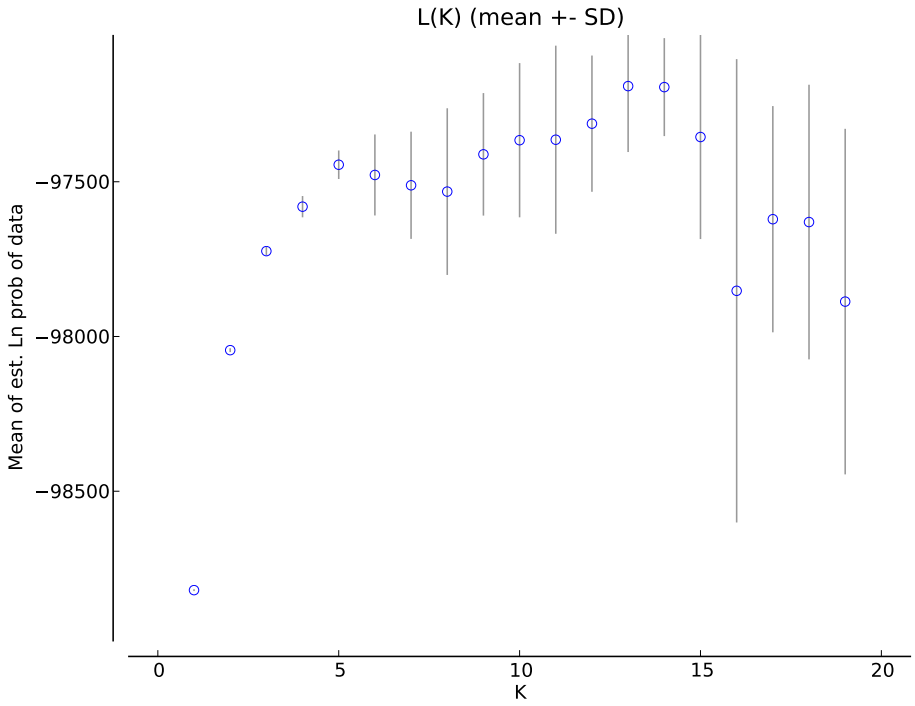


Figure A6: The log-likelihood from STRUCTURE from $K = 2$ to 19.

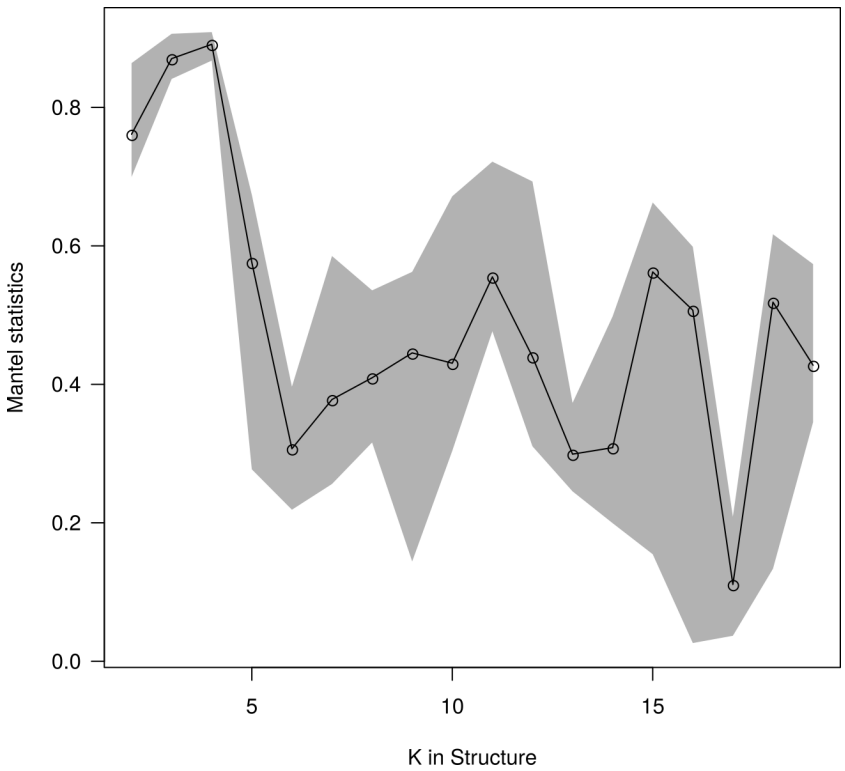


Figure A7: Mantel test statistic and its 95% percentile interval from a randomization procedure between distance matrices from AFM and STRUCTURE from $K = 2$ to 19.

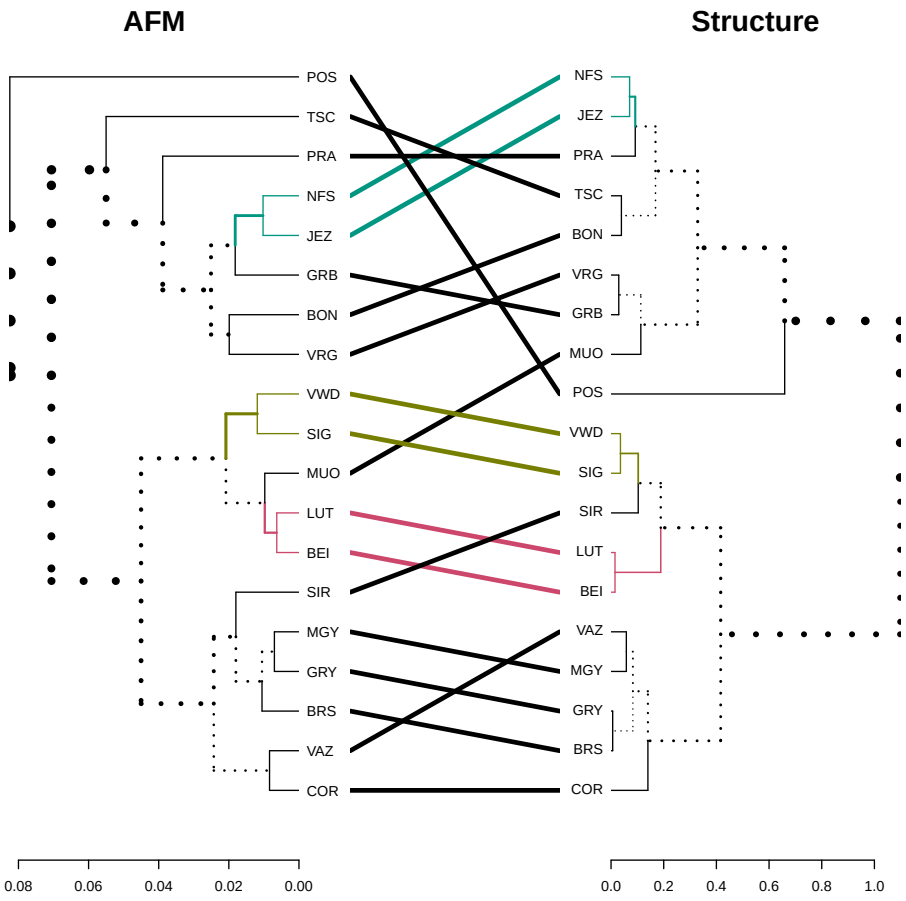


Figure A8: Entangled dendrograms from AFM and STRUCTURE with $K = 4$. Colors highlight populations that belong to the same clusters.

Supplementary Methods S3: Details of the modifications to the methods proposed by Ovaskainen *et al.* (2011) and Karhunen *et al.* (2014)

Following the notation of Ovaskainen *et al.* (2011), if we consider only additive effects, the vector of additive values of all traits for individual i is $\mathbf{a}_i$, and the matrix of additive vectors for all individuals is $\mathbf{A} = (\mathbf{a}_i)_i$. Then, the mean additive value of population X, $\mathbf{a}_X^P$, can be obtained as the mean of the additive values of all individuals in population X, and the matrix of additive vectors for all populations is $\mathbf{A}^P = (\mathbf{a}_X^P)_X$. When populations are derived from a common ancestral population and the trait values are normally distributed, under drift, the matrix of population-level effects, $\mathbf{A}^P$, is expected to follow the multivariate normal distribution as

$$\mathbf{A}^P \sim N(\boldsymbol{\mu} \otimes \mathbf{I}_{n_P}, 2\mathbf{G}^A \otimes \boldsymbol{\theta}^P), \tag{1}$$

where $\boldsymbol{\mu}$ is the vector of expected additive trait means determined by the allele frequencies in the ancestral population, n_P is the number of populations and $\mathbf{I}_{n_P}$ is an $n_P \times n_P$ identity matrix, $\mathbf{G}^A$ is the ancestral variance-covariance matrix, $\boldsymbol{\theta}^P$ is the population-to-population coancestry matrix, and $\otimes$ is the Kroenecker product. $\boldsymbol{\theta}^P$ can be estimated assuming the admixture F-model, while $\boldsymbol{\mu}$, $\mathbf{A}^P$, and $\mathbf{G}^A$ can be co-estimated using the Bayesian mixed-effects animal model accounting for the family structure of the common garden (i.e. the pedigree) and $\boldsymbol{\theta}^P$ (Ovaskainen *et al.*, 2011). Then, the additive genetic variance-covariance matrix of the contemporary populations assuming no selection, $\mathbf{G}$, can be estimated as $\mathbf{G} = 2\mathbf{G}^A(1 - \theta^S)$, where θ^S is the mean within-population (or self) coancestry of all populations, thus the drift distance of the contemporary populations from the ancestral population. Here, we estimated the heritability of traits and the genetic correlation between trait pairs as the proportion of the observed phenotypic variance and covariances that are additive (i.e. using $\mathbf{G}$) (Falconer & Mackay, 1996). The evidence for selection can be summarized using the S-statistic calculated as the Mahalanobis distance between $\mathbf{A}^P$ and the distribution of equation 1. $S = 0.5$ indicates consistency with neutrality, $S = 0$ implies a match with purifying, and $S = 1$ with diversifying selection. Thus, S measures the overall signature of selection across all populations.

In this study, we assess to what extent the particular populations deviate from their neutral expectation. Population X, whose 95% of the posterior distribution of $\mathbf{a}_X^P$ is outside of the neutral envelop defined as $\boldsymbol{\mu} \pm \sqrt{2\mathbf{G}^A \boldsymbol{\theta}_X^S}$ strongly contributes to the overall selection signal captured by the S statistic. The neutral envelop can be uni- or multi-variate depending if one or multiple traits are studied. Our motivation for this population-wise evaluation of divergence is that a single S-statistic cannot distinguish between the two scenarios of many populations that are slightly diverged or a single/few populations that are diverged to a great extent.

The H-statistics measure if the distance between the populations' mean additive trait values is more similar to the environmental distances than expected based on drift (Karhunen *et al.*, 2014). The original H-test is based on the Mantel test statistic, i.e. the product moment between the distance matrices of environment and traits. As the covariance is not only influenced by the correlation between the matrices but also by the absolute values of trait differences, the original H-test may yield false positive results. In particular, in cases with strong evidence for selection, the H statistic can be high (i.e. close to 1) even when selection is uncorrelated with the tested environmental driver (Fig. A9). For this reason, we propose the use of the H^* statistic, which is the Pearson or standardized Mantel statistic,

thus the H^*-test can be viewed as a standardized version of the H-test. We note that the superiority of a standardized Mantel statistic has already been pointed out in the context of spatial distance matrices (Legendre & Fortin, 2010).

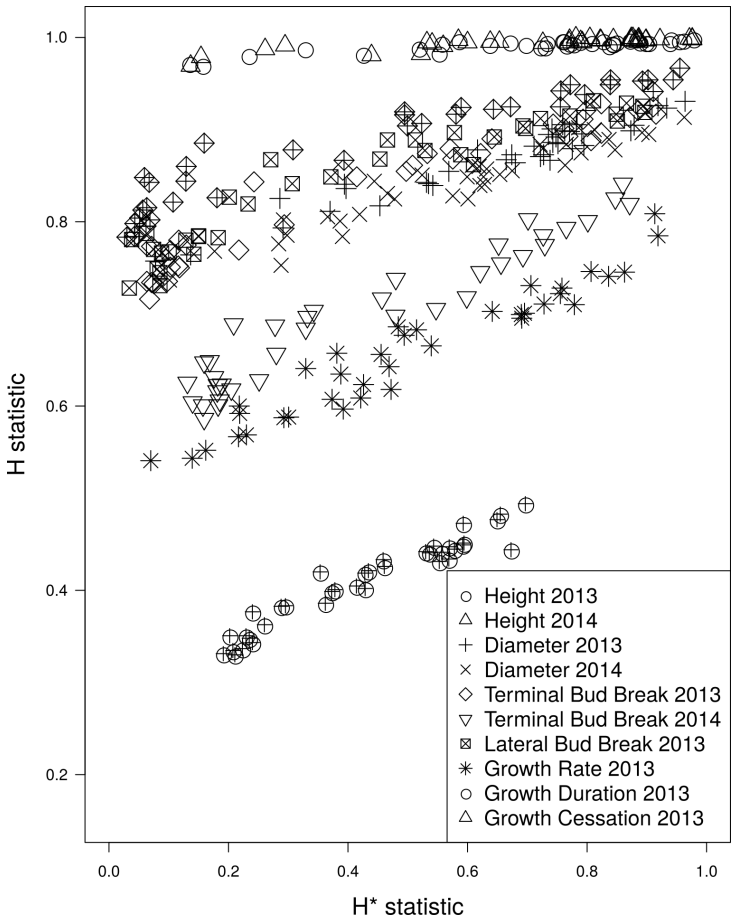


Figure A9: Comparison of the H- and H^*-statistics for the 10 studied traits. Each point indicates the H- and H^*-statistics for one of the environmental variable listed in Table 1 of the main text.

References

Burga CA, Hussendörfer E (2001) Vegetation history of *Abies alba* Mill. (silver fir) in Switzerland–pollen analytical and genetic surveys related to aspects of vegetation history of *Picea abies* (L.) H. Karsten (Norway spruce). *Veg Hist Archaeobot*, **10**: 151–159.

Earl DA, *et al.* (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*, **4**: 359–361.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol*, **14**: 2611–2620.

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. Longmans Green, Harlow, Essex, UK, 4 edn.

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.

Frank A, Sperisen C, Howe GT, Brang P, Walthert L, Clair JBS, Heiri C (2017) Distinct genecological patterns in seedlings of Norway spruce and silver fir from a mountainous landscape. *Ecology*, **98**: 211–227.

Galili T (2015) dendextend: an r package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics*, **31**: 3718–3720.

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci*, **7**: 457–511.

Geweke J (1991) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: *Bayesian Statistics 4* (eds. Bernado J, Berger J, Dawid A, Smith A). Clarendon Press, Oxford, UK.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *J Stat Softw*, **22**: 1–19.

Hansen TF, Pélabon C, Houle D (2011) Heritability is not evolvability. *Evol Biol*, **38**: 258.

Heidelberger P, Welch PD (1981) A spectral method for confidence interval generation and run length control in simulations. *Commun ACM*, **24**: 233–245.

Houle D (1992) Comparing evolvability and variability of quantitative traits. *Genetics*, **130**: 195–204.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour*, **9**: 1322–1332.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**: 1801–1806.

Karhunen M, Ovaskainen O (2012) Estimating population-level coancestry coefficients by an admixture F model. *Genetics*, **192**: 609–617.

Karhunen M, Ovaskainen O, Herczeg G, Merilä J (2014) Bringing habitat information into statistical tests of local adaptation in quantitative traits: A case study of nine-spined sticklebacks. *Evolution*, **68**: 559–568.

Legendre P, Fortin MJ (2010) Comparison of the mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Mol Ecol Res*, **10**: 831–844.

Ovaskainen O, Karhunen M, Zheng C, Arias JMC, Merilä J (2011) A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. *Genetics*, **189**: 621–632.

Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**: 945–959.