Adaptive genomic variation associated with environmental gradients along a latitudinal cline in *Rana temporaria*

Authors: Alexandra Jansen van Rensburg[a,b], Maria Cortazar-Chinarro[c], Annsi Laurila[c], and Josh Van Buskirk[a,d]

[a] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Switzerland

[b] Present address: School of Biological Sciences, University of Bristol, UK; ORCHID ID: 000-0002-5093-7040

[c] Department of Ecology and Genetics, Uppsala University, Sweden

[a] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Switzerland

Correspondence: A. Jansen van Rensburg
School of Biological Sciences
Life Sciences Building
24 Tyndall Avenue
Bristol BS1 8TQ, UK
email: alexjvr@gmail.com

Abstract

Rana temporaria occur across a large geographic and environmental gradient in Scandinavia. Several studies involving common garden experiments have established adaptive divergence across the gradient. The main objective of this study was to determine the extent of neutral and adaptive genetic divergence across the latitudinal gradient. Here we sequence genome-wide markers for 15 populations from six regions sampled from southern Sweden to Finland. Using a multivariate approach we find that 68% of the genomic variation is associated with climate or geographically structured climate. Using outlier scans and environmental association analyses we identify a set of potentially adaptive loci and examine their change in allele frequency associated with different climatic variables. Using a gradient forest analysis we identify points along three of the climate variables where allele frequencies change more rapidly than expected if it were a linear association. We identify a large threshold effect associated with BIO5 (mean temperature during the warmest month) which is seen as a rapid change in southern Sweden. By comparing the change in neutral and adaptive allele frequencies across the whole gradient, we identify southern Sweden as a region with the largest divergence between the datasets. This suggests small changes in the climate may result in a mismatch between the adaptive genotypes and the environment in these populations. Overall this study shows that genomic analyses can provide a powerful complement to common garden experiments to improve our understanding of adaptive divergence across heterogeneous landscapes.

Introduction

The geographic distribution of genetic variation across a species range is an important determinant of population persistence under changing environmental conditions (Hoffmann & Sgrò 2011). To mitigate future biodiversity loss due to climate change it is important that we identify the most important environmental drivers of adaptive divergence and determine how genetic variation contributes to adaptation. Landscape genomics methods have been employed to identify potentially adaptive loci across many study systems (reviewed in Rellstab *et al.* 2015). However, adaptation across environmental gradients are often characterised by divergence in multiple phenotypic traits with polygenic genetic underpinnings (Pritchard & Di Rienzo 2010; Yeaman 2015). This presents a challenge when examining non-model organisms, where the lack of genomic resources is often prohibitive in determining the genomic architecture and genetic basis of adaptation (Manel *et al.* 2016). However, by developing our understanding of the distribution of genetic variation (adaptive and neutral) in geographic and climate space, we can draw meaningful conclusions about the ecological determinants of species distributions and adaptive divergence without pinpointing the underlying causal variants (Jones *et al.* 2013; Fitzpatrick & Keller 2015; Forester *et al.* 2016).

Two main statistical models have been developed to identify potentially adaptive loci while accounting for population structure (reviewed in Hoban *et al.* 2016). Population genetic methods identify loci with higher than expected differentiation (usually measured by F_{ST}) compared to neutral expectations based on population structure (Luikart *et al.* 2003). These methods are effective for detecting selective sweeps associated with strong selection acting on a few beneficial loci. However, the signal of divergence is difficult to detect if there is high gene flow between populations adapted to different conditions, or when the adaptive traits are polygenic with many small changes in allele frequency additively contributing to adaptation (Kawecki & Ebert 2004; Pritchard & Di Rienzo 2010). The second kind of model, termed Environmental Association Analysis (EAA; reviewed in Rellstab *et al.* 2015), is aimed at finding associations between allele frequencies and environmental

variables, thus does not rely on strong sweeps underlying adaptation. This approach is particularly useful for comparing the importance of different environmental variables and for mapping spatial changes in allele frequencies of adaptive loci. However, EAA assumes a linear relationship between allele frequencies and environmental gradients (Thomassen *et al.* 2010; Fitzpatrick *et al.* 2015), thus confining inferences to linear responses. But non-linear patterns of genetic variation along environmental gradients are probably common, judging from laboratory studies of organismal physiology (Angilletta's 2009 thermal adaptation book, for example), and could be important for identifying populations or geographic regions that are especially vulnerable to climate change. Thus, it is important to modify EAA such that non-linear relationships between allele frequencies and environmental gradients can be detected in nature. Here we combine a modelling approach to loci identified using FST outlier and EAA methods to identify such non-linear associations with environment in the European common frog, *Rana temporaria* across a latitudinal gradient.

Rana temporaria is widespread across Europe and occur throughout Scandinavia (Sillero *et al.* 2014). Populations occur across a wide range of habitats, which suggests adaptive divergence across the species range. Common garden experiments across 1600-km of the Scandinavian latitudinal gradient have confirmed this by establishing extensive latitudinal variation in larval and adult life history traits (Merila *et al.* 2000; Laugen *et al.* 2003b, 2005a; Palo *et al.* 2003a; Lindgren & Laurila 2005). The geographic scale of the gradient provides an interesting system to investigate the genomics of adaptation to environment, because there is little gene flow between populations adapted to different environments.

The main objective of this study is to determine the extent of neutral and adaptive genetic divergence across the latitudinal gradient. Specifically, we aim to 1) characterise the neutral genetic structure, 2) determine the proportion of the genome associated with environmentally driven adaptive divergence, and 3) identify environmental thresholds to adaptation by examining the non-linear response of adaptive loci to climate variables. We were particularly interested in determining whether there are non-linear relationships between adaptive allele frequencies and environmental variables. Such a response would suggest that there is a threshold

along that particular environmental variable that requires a larger change in allele frequency than would be expected based on the gradual change in that environmental variable. The results have important implications for identifying populations and geographic regions that would be particularly vulnerable to changing environments.

Methods

Sampling & DNA extraction - To determine intra-specific population structure and adaptive variation, 163 individuals were collected from six geographic regions across ~1500 km of the Scandinavian latitudinal gradient (Fig. 1; Table 1). Each region was represented by one to three populations (where a population is a pond), for a total of 15 populations in the final dataset. At each sampling site, we sampled approximately 10 eggs from 20-30 freshly –laid clutches (less than two days old). The eggs were transported to the laboratory at Uppsala University where they were raised in separate containers kept in climate room at 16 °C. Tadpoles were raised to Gosner stage 25 (Gosner 1960), whereafter they were euthanized with an overdose of MS222, preserved in 96% ethanol and stored at 4 °C until DNA extraction. Total DNA was extracted from one individual per clutch (henceforth family) using the Qiagen DNeay blood and tissue kit (Qiagen, CA, USA).

ddRAD sequencing, de novo assembly, and variant calling - To establish a genome-wide marker set, double digest restriction-site-associated DNA sequencing libraries were prepared with the restriction enzymes *EcoRI* and *MseI*, using a modification of the protocol by Peterson *et al.* (2012). We constructed 6 libraries comprising 48 samples each for single-end sequencing (125bp) on an Illumina HiSeq 2500 v4 at the Functional Genomics Center, University of Zurich. Individual samples were identified by unique 5-bp barcodes. Raw sequence reads were demultiplexed using the process_radtags package from Stacks (Catchen *et al.* 2011). Demultiplexing was based on a unique 9-bp sequence for each individual (5-bp

unique barcode + 4-bp restriction enzyme recognition site), with 1 mismatch allowed. Adapter and other Illumina-specific sequences were removed using Trimmomatic v0.33 (Bolger *et al.* 2014).

De novo assembly and variant calling was implemented using pyRAD (Eaton 2014) which first finds clusters within individuals based on a clustering threshold and minimum depth, and then clusters these loci between individuals (using the same clustering threshold), and identifies loci found in a user specified number of individuals. Clustering thresholds of 90-99% sequence similarity were tested; an optimum of 94% was chosen because it maximised nucleotide diversity and minimised the estimated number of paralogs in the dataset. A maximum of 4 sites with a Phred quality score <20 were allowed per sequence. Clusters were kept if they had >5x coverage per individual and were found in at least 4 individuals.

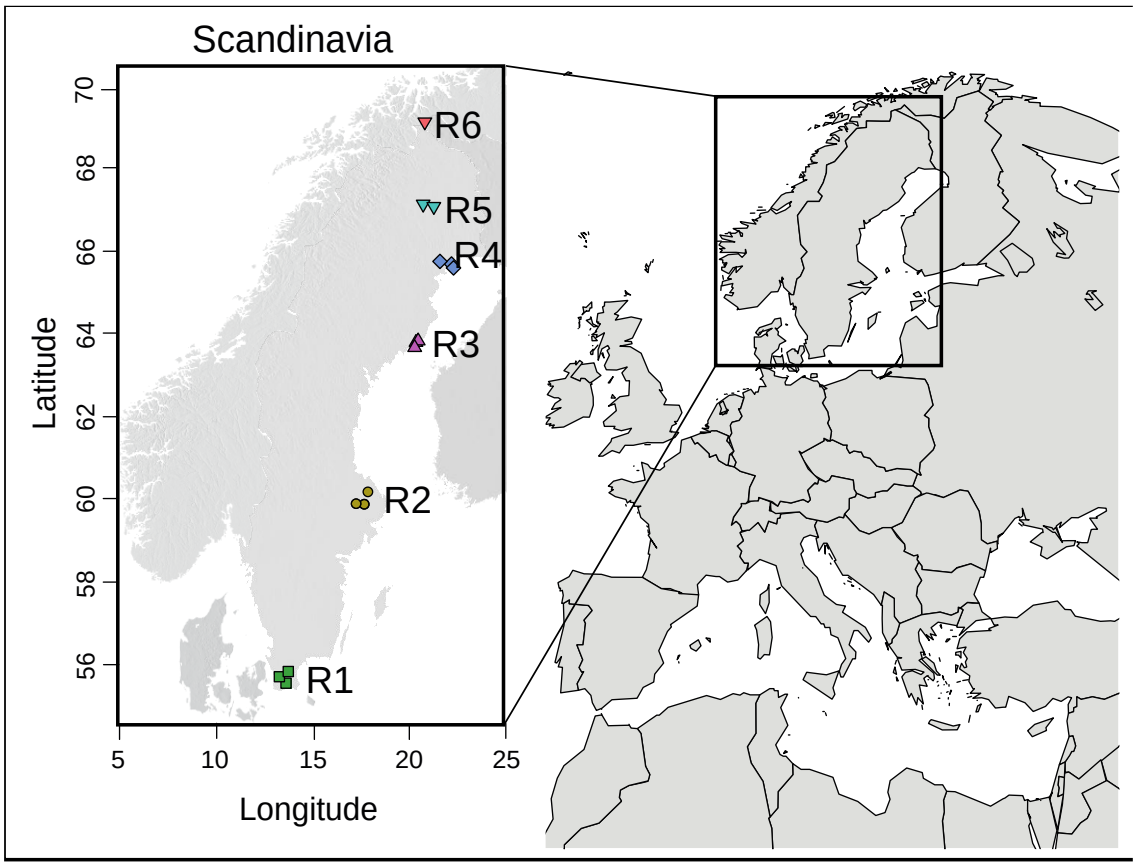


Figure 1 Sampled populations across the Fennoscandian latitudinal gradient. One to three populations were sampled within each of six regions. Symbols represent the five genetic clusters found at K=5, the most likely number of divisions found by

161 DAPC, PCAdapt, and SNMF analyses. Geographic regions are named R1 - R6 from
162 south to north along the latitudinal gradient.

Table 1 Summary of the diversity statistics calculated per population. The geographic coordinates (Lat, Long) are shown for each population (Pop). The geographic region (Region) to which the populations belong is shown along with the abbreviation (R1-R6) used throughout the manuscript. The number of individuals included in the analyses (n) are shown. Gene diversities (Hs), deviation from random random mating (F_{IS}), observed heterozygosity (Ho), and the average number of heterozygous sites across all sequenced sites (AvgHet) with standard deviation (SD) are reported per population.

Region	Pop	Lat	Long	n	F_{IS}	Hs	Ho	AvgHet	SD
All				132	0.10	0.23	0.21	0.0025	
	Sk.Ho	55.859	13.764	9	-0.01	0.29	0.29	0.0033	0.0004
R1: Skåne ■	Sk.SF	55.558	13.638	10	0.12	0.27	**0.23**	0.0030	0.0001
	Sk.SL	55.723	13.287	17	0.14	0.27	**0.23**	0.0031	0.0001
	Upp.Gra	59.878	17.667	9	0.07	0.19	**0.18**	0.0023	0.0001
R2: Uppsala ●	Upp.K	59.891	17.242	10	0.08	0.22	**0.20**	0.0024	0.0001
	Upp.O	60.178	17.854	9	0.07	0.21	**0.19**	0.0024	0.0001
	Um.Gr	63.792	20.367	9	0.07	0.21	0.19	0.0023	0.0001
R3: Umeå ▲	Um.Taf	63.830	20.486	10	0.09	0.22	0.20	0.0024	0.0001
	Um.UT3	63.658	20.298	2	0.01	0.20	0.18	0.0021	0.0003
	LT1	65.684	22.213	9	0.06	0.21	0.20	0.0024	0.0001
R4: Luleå ◆	LT2	65.750	21.602	3	0.03	0.24	0.21	0.0024	0.0001
	LT3	65.583	22.319	7	0.05	0.22	0.20	0.0024	0.0001
R5: Kiruna ▼	Kir.G	67.111	20.656	10	0.09	0.23	0.20	0.0025	0.0001
	Kir.L	67.052	21.224	10	0.08	0.22	0.20	0.0025	0.0001
R6: Finland ▼	FIN	69.044	20.805	8	0.10	0.22	0.19	0.0023	0.0002

SNP validation - The putative variants identified though the pyRAD pipeline were filtered for possible sequencing errors, paralogs, and uninformative SNPs. The following filters were applied: 1) SNPs that were genotyped in less than 50% of individuals were removed using the *--max-missing* function in VCFtools v0.1.12b (Danecek *et al.* 2011). 2) Loci with a minor allele frequency less than 0.05 in the full dataset were removed as they are more likely to be sequencing error, and if they are true variants they are uninformative and likely to bias tests for selection (Roesti *et al.* 2012). 3) Using a sliding-window of 10-bp we tested whether the number of variants increased towards the end of the sequence. No significant difference was found between bins, thus the sequences were not trimmed further. 4) We reduced linkage in the dataset we included only one variant per locus using the *--thin* function in VCFtools. 5) We assessed whether loci were in Hardy-Weinberg equilibrium (HWE) within each population using the *--hardy* function in *PLINK* (Purcell *et al.* 2007). Loci with an observed heterozygosity more than 0.5, and loci that deviated significantly from HWE based on the exact test ($p<0.05$; (Wigginton *et al.* 2005) were removed from the dataset if they occurred in more than 5 populations. 6) We then calculated linkage disequilibrium for each locus pair per population in PLINK v 1.07 (Purcell *et al.* 2007). Loci with $r^2 > 0.8$ never occurred in more than 5 populations, so no loci were excluded at this step. 7) Finally we excluded individuals with more than 55% missingness.

Summary statistics - Nucleotide diversity was calculated for each sample as the frequency of heterozygous bases (IUPAC codes) from the *pyRAD* output, and means were calculated per population. Although these calculations are based on sequences before final filters are implemented, two reasons convince us that these results are robust: 1) *pyRAD* calculates a binomial probability that a base is homozygous or heterozygous based on a maximum likelihood approach of jointly estimating the heterozygosity and sequencing error rate from the base frequencies within each individual's stacks. If the read depth of the stack falls below a user-set threshold, or is too low to make a statistical base call, the base remains undetermined. Thus the final base calls per individual should be fairly robust (Eaton 2014). 2) The post-*pyRAD* filters outlined in the previous paragraph do not address sequencing or SNP-calling

errors, but rather minimises missingness, systematic biases, and linkage between loci.

Further summary statistics were calculated in R v3.3.1 (R core team 2016) with the *hierfstat* v*0.04-22* and *adegenet v2.0.1* packages (Goudet 2005; Jombart 2008; Jombart & Ahmed 2011). Gene diversities (Hs), deviation from random mating (F_{IS}), and observed and expected heterozygosity are reported per population (Table 1).

Population structure - Pairwise population differentiation was estimated using Weir and Cockerham's F_{ST} (Weir & Cockerham 1984) as implemented in *hierfstat* (Goudet 2005). We visualised the genetic distance between populations with a principal component analysis (PCA) implemented in *PCAdapt v3.0.3* (Luu *et al.* 2016). The following analyses were conducted in *adegenet.* To test for isolation by distance, we calculated the correlation between log transformed pairwise geographic distances and scaled pairwise genetic distances (F_{ST}/1-F_{ST}) (Rousset 1997). We tested for significance using a Mantel test. We quantified the proportion of genetic variation that explained differentiation within and between populations with an analysis of molecular variance (AMOVA). Finally, we performed a discriminant analysis of principal components (DAPC) to determine the most likely number of clusters in the dataset and visualise broad-scale population structure.

Climate Data - We obtained climate data from WorldClim v1.4 (Hijmans *et al.* 2005) at a resolution of 2.5 minutes of degrees using the R package *raster v.2.5-8* (Hijmans *et al.* 2015). Many of these variables are derived from the same data and are highly correlated. To reduce the redundancy in the climate variables retained for analyses, we first calculated correlation between all variables, and then removed variables if they exceeded a correlation threshold. We calculated the absolute values of pairwise ranked correlation (Spearman's rho) between all 19 BioClim variables from WorldClim, longitude, latitude, and season length. Season length was calculated as the number of days above 6 and 8 °C at each sampling site, since 6 °C is approximately the development threshold of *R. temporaria* tadpoles (Laugen *et al.* 2003a). We reduced redundancy in the environmental dataset by detecting pairs of

variables that had an absolute correlation >0.8, and then eliminating the one that had the highest mean correlation with all other variables (Kuhn et al. 2016). This procedure retained five BioClim variables, and these were used as the environmental variables in all remaining analyses (Fig. 2).

Relative contribution of environment and IBD to genomic differentiation - We investigated the effects of climate and geography on neutral genetic structure using full and partial redundancy analyses (RDA) with variance partitioning. Redundancy analysis is a form of multivariate regression, which can be used when both the predictor and response variables are multivariate (Legendre & Legendre 2012). As a canonical extension of multiple linear regression, RDA identifies a set of orthogonal linear predictor variables that explains the most variation in a set of linear response variables. In this case each RDA axis represents a set of co-varying loci (response variables), which are correlated with co-varying environmental variables (predictor variables). RDA has greater power to detect multivariate genotype-environment relationships than methods based on distance matrices or Mantel tests (Legendre & Fortin 2010).

We created two matrices as response variables: 1) the 5 climate variables identified above, centered and standardised, and 2) geographic coordinates of each sampling site. The response matrix was the minor allele frequencies of 2081 loci for each individual. We ran a sequence of nested models to partition variation in climate and geography as explanatory variables of allele frequencies: (1) the full model including both climate and geography; (2) climate only, with the influence of geography partialled out (climate | geography); and (3) geography only, with climate partialled out (geography | climate). The difference in the variance explained by model (1) minus the sum of models (2) and (3) was interpreted as the contribution of climate and geography acting together. Overall and residual variance was calculated for each model, and the model significance was tested with 999 permutations. The RDA was conducted using the R package *vegan v2.4-1* (Oksanen *et al.* 2015).

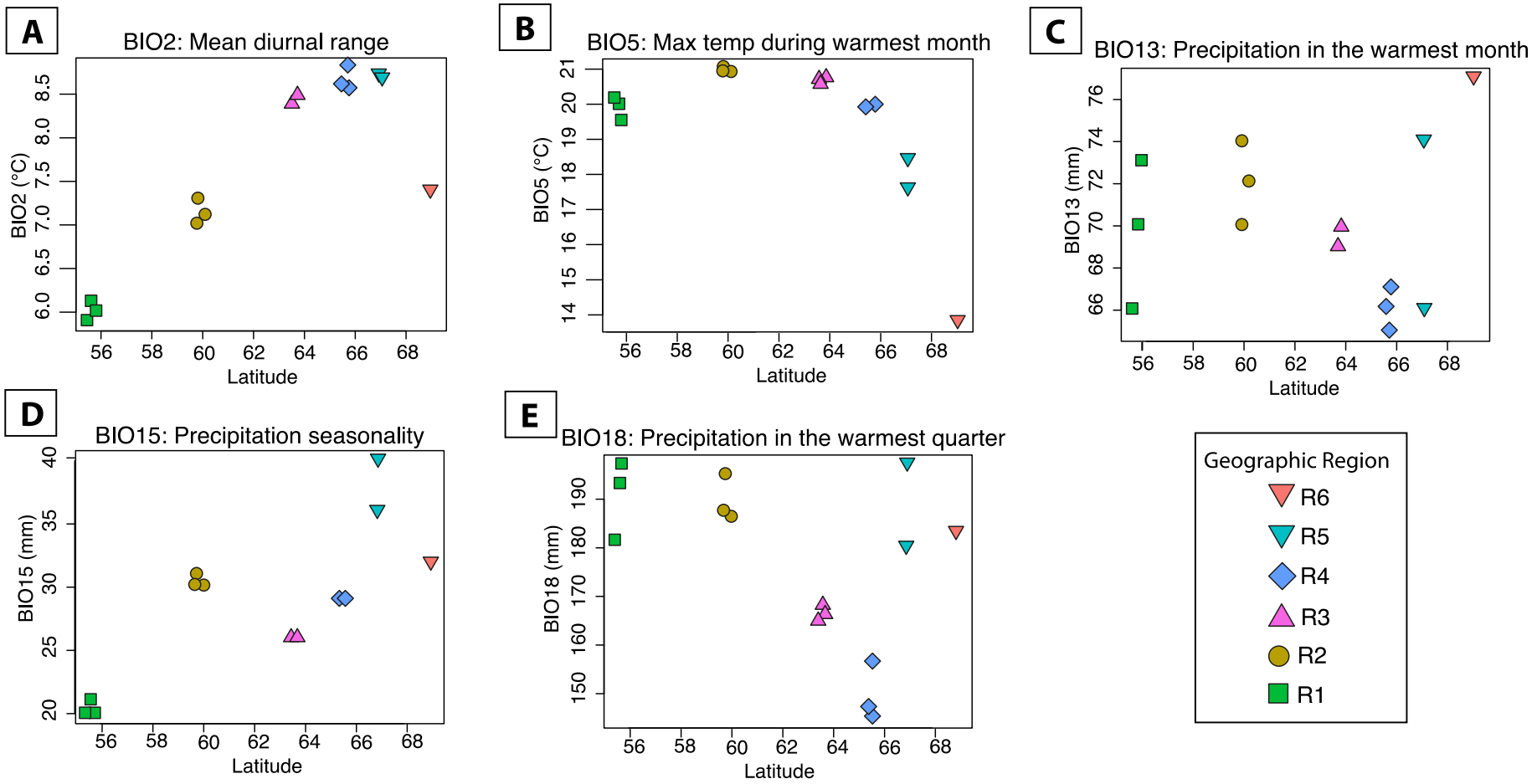


Figure 2 Latitudinal distribution of the five BioClim variables used in this study. Each point represents a sampled population.

Signatures of adaptation: Outlier detection and genotype-environment associations - We created two datasets containing loci potentially under selection using two common approaches: 1) Outlier analyses to identify loci that are more differentiated between populations than expected under a neutral model, and 2) Environmental Association Analyses (EAA) to identify loci strongly associated with an environmental variable (reviewed in Rellstab *et al.* 2015). These approaches are likely to identify different loci, since their underlying assumptions are different. The Outlier dataset comprised loci identified with *PCAdapt* (Luu *et al.* 2016), and from the X_TX statistic calculated in *bayenv2* (Günther & Coop 2013). The EAA dataset comprised loci identified using *bayenv2* and LFMM (Frichot *et al.* 2013).

PCAdapt identifies outlier loci as those that are more associated with population structure than expected. We used the R package *pcadapt 3.0.4* (Luu *et al.* 2016), which calculates a vector of z-scores of the how related each SNP is to the first K principal components, where K is the user-specified number of population clusters. A Mahalanobis distance is then calculated for each SNP to determine whether it deviates from the main distribution of z-scores. These scores are scaled by a constant, the genomic inflation factor, which produces a chi-squared distribution of values with K degrees of freedom. K was calculated as the most likely number of genetic clusters after testing K 1-20 and inspecting the scree plot of the proportion of explained variance for each K (Fig. S1). Based on these results, we chose K=5 for further analyses. We used a false discovery rate of 10% to identify outlier loci.

Bayenv2 estimates genotype-environment associations and an F_{ST}-like statistic (X_TX) while correcting for covariance of allele frequency between populations due to neutral processes. We used *bayenv2* to identify loci for both datasets. First we estimated the neutral covariance matrix based on 500 randomly selected loci. Two independent runs with 100000 MCMC iterations were run. We tested for convergence within each run by calculating Pearson's product-moment correlation (*cor.test* in R) between the final matrix and nine matrices printed out at 10000 step intervals (9 correlations). We constructed distance-based trees to determine whether relationships among populations remained constant within and between runs. Convergence between runs was calculated as the correlation between the final matrix in each run. Since these were highly correlated, we arbitrarily chose the final

matrix from the first run as our final covariance matrix. The full model was run using this covariance matrix, a file containing standardised measures of each environmental variable, and a genotype file containing SNP counts across all populations. We ran a non-parametric test that calculates the Bayes factor, Spearman's *p,* and Pearson's correlation coefficient for each genotype-environment association (-t -c -r). In addition we calculated the X_TX population differentiation statistic (-X). For this test, Gunther *et al.* (2013) suggest ranking loci by their X_TX statistic rather than selecting those above a specific threshold.

We conducted three independent runs with *bayenv2* of 100,000 MCMC iterations each for each of the five genotype-environment associations, and tested convergence by calculating the correlation between runs for each statistic (BF, *p*, X_TX). We also compared the overlap in loci identified in the top 5%, 6-10%, and 11-15% ranked loci based on the X_TX statistic for each environmental variable to ensure the repeatability of the results. We then ran an additional 7 independent *bayenv2* runs, and calculated the median result across all 10 runs as our final output. We selected the top 100 ranked loci based on the X_TX statistic for the outlier dataset. For the EAA dataset we selected loci with a log10 Bayes Factor (BF) >0.5 (Kass & Raftery 1995), and absolute Spearman's rho (*p*) >0.3.

Finally, we screened the dataset for additional EAA loci using the Latent Factor Mixed-effect Model (LFMM; Frichot *et al.* 2013; Frichot & François 2015). LFMM calculates the correlation between genotype and environment while simultaneously accounting for population structure with latent factors incorporated in the model. The number of latent factors is user specified, and should represent the number of genetic clusters (K) that best describes the population structure in the dataset. As suggested by Frichot & François (2015), we estimated the most likely K by evaluating the cross-entropy criterion for K1-10 using the function *snmf* in the R package *LEA*. The most likely K was 5, which is consistent with the population structure analyses described above, and therefore LFMM was run with K=5 for each of the five environmental variables. The Gibbs sampler algorithm was run five times for each environmental variable, with 10000 cycles and a burn-in of 5000 cycles. The median of the resulting correlation scores (z-scores) was calculated across all five runs. The authors suggest a recalibration of the mean z-scores by lambda; that is the

square of the mean z-scores divided by the median of a chi-squared distribution with one degree of freedom (λ; ~0.455). Lambda should be close to one − but more importantly, this should produce the correct adjusted p-value frequency distribution. Adjustment of λ can correct for liberal or conservative p-value distributions. We evaluated the effect of λ = 0.45-1.00 on the p-value distribution for each of the 5 environment-genotype associations. The shape of the distribution did not change much, but the frequency of p-values >0.1 increased as λ increased (i.e. the correction was more conservative). Thus, for a lambda close to one and the correct adjusted p-value distribution, we chose λ = 0.85 for BIO2, BIO5, BIO15, and BIO18, and λ = 0.45 for BIO13. To control for false discoveries, we applied a Benjamin-Hochberg adjustment with a false discovery rate of 5%.

Genomic Turnover across ecological gradients - We assessed how genomic variation changes across Scandinavia and whether important climatic thresholds occur by fitting a Gradient Forest model (Ellis *et al.* 2012) to each of the SNP datasets. We compare the change in allele frequency between the adaptive loci (Outlier and EAA dataset) and a Reference dataset composed of all the remaining loci. The Gradient forest model was developed as an extension of the random forest model to assess community level responses to ecological gradients, and has recently been applied to genomic data to detect non-linear change in allele frequencies along ecological gradients (Fitzpatrick & Keller 2015). It is a machine-learning ensemble approach that fits multiple regression trees between allele frequency and environmental variables. A set of decision trees is built to describe change in allele frequency across the predictor variable range. Each split is determined by minimising the "impurity" in the data, i.e. minimising the sums of squares of the allele frequency and thus maximising the tree fit. This means the split will always describe the biggest change in allele frequency at the current point in the tree, and a relative split importance can be calculated.

The response variable was the minor allele frequencies (MAF) for each SNP dataset. We included only loci variable in more than 4 populations to ensure robust regressions. The predictor variables were the 5 BioClim variables described above. To account for unsampled geographic structure in the dataset, we also included

Moran's Eigenvector Map variables (MEM; Dray et al. 2006), which are orthogonal vectors maximising spatial autocorrelation between sampled locations. Broad-scale spatial structure is most likely explained by the most positive eigenvectors (Manel et al. 2010b, 2012; Sork et al. 2013), so we included the first half of the positive MEMs here; in total three MEM vectors.

The gradient forest model was fit to each SNP dataset using the R package *gradientForest* (Ellis *et al.* 2012). We constructed 2000 regression trees per SNP, with default values for the variable correlation threshold (0.5), the number of candidate predictor variables sampled at each split (2), and the proportion of samples used for training (~66%) and testing (~33%) each tree. The relative importance (R^2) of each predictor variable was calculated as the weighted mean of the proportion of variance explained by the validation data. The cumulative importance for the change in allele frequency for each locus was calculated as the sum of the split importance across climatic variables, and the mean allelic turnover per climatic variable was calculated for each of the three SNP datasets.

Changes in allele frequency across the landscape were visualised by transforming each climatic variable by the genomic importance calculated for each SNP dataset; i.e. we produced a transformed dataset for the Reference, EAA, and Outlier datasets. The three transformed datasets were produced for climate data extracted from a raster stack covering eastern Sweden and northernmost Finland. For each dataset, the transformed variables were reduced by PCA, and a colour from the RGB colour palette in R was assigned to each of the first three principal components. Thus, for each geographic point along the latitudinal gradient, a single colour was used to represent the genomic composition in three-dimensional principal component space. To compare the genomic turnover between the neutral and two candidate SNP datasets, we calculated the distance in genomic space at each geographic point as the Procrustes residuals between the pairs of transformed matrices calculated above. The genomic difference between datasets was normalised and mapped in geographic space as before.

Results

ddRAD data generation and variant filtering - The final dataset consisted of 132 individuals from 15 populations (2-17 individuals per population; Table 1). There were 2081 SNP loci, with a mean depth of 17.9 - 25.5x and genotyping rate of 72-97% per population. A summary of the number of raw reads, the output from pyRAD, and the final dataset can be found in Table S1.

Genetic variation and Population structure - The highest genetic diversity was found in the southernmost region, R1, while the rest of the gradient was characterised by lower genetic diversity that was similar across all regions (Table 1). Specifically, the mean gene diversity (H_S) and heterozygosity (H_E) in Skåne were 0.28 and 0.25, respectively, while they were only 0.22 and 0.19 across the rest of the gradient. Similarly, the frequency of heterozygous sites across all sequenced sites averaged ~1/300 (0.0031) in R1, and ~1/400 (0.0024) across the rest of the gradient.

Measures of genetic differentiation showed evidence of population structure within and between regions. AMOVA indicated that most genetic variation (68.72%) was found within populations, while significant variation was found among populations within regions (6.05%) and among (25.23%) regions (Table S2). The mean global F_{ST} was 0.21 (Fig. 3, Table S3), suggesting strong population structure on average between populations. However, pairwise genetic distance was much lower within (F_{ST} = 0.06) than between (F_{ST} = 0.16) regions, and there was significant isolation by distance (R=0.434, p=0.001) across the sampled area.

To determine the broad-scale population structure, we first visualised the genetic distance between populations using PCA, and then estimated the most likely number of genetic clusters using DAPC. The first two axes of the PCA explained approximately 24% of the variance. PC1 (~15%) partitioned the two southern regions (R1 and R2) from the rest, and PC2 (~9%) partitioned populations latitudinally, with a graded differentiation from R2 to R6 (Fig. 4).

Discriminant analysis of principal components (DAPC) predicted five genetic clusters, corresponding to R1, R2, R3, R4, and the three northern populations (Fig. 1). When separating the dataset into increasing numbers of clusters, populations

grouped out sequentially from south to north, except that R5 and R6 always grouped together.

Redundancy Analysis - The full RDA model included climate and geographic coordinates and explained 76.6% of the total genetic variation (*p=0.001*). Based on the partial RDA, climate was significantly associated with genetic variation (climate | geography; *p=0.002*), and explained 49.5% of the total variation. The variation explained by geography alone (geography | climate) was much less (11.4%) and was non-significant. The proportion of genetic variation explained by spatially structured climate (climate ∩ geography) was 39.1%.

Signature of adaptation: Outlier detection and genotype-environment associations - *PCAdapt* identified 50 outlier SNP loci and the X_TX statistic from *bayenv2* returned the top 100 outlier loci. A total of 28 loci were identified by both methods, so that the Outlier dataset comprised 122 unique loci (Fig. S2).

The EAA dataset comprised loci identified using *Bayenv2* and LFMM as associated with the five chosen BioClim variables. *Bayenv2* identified 123 unique loci (6% of the total loci tested), with 13% of these loci associated with multiple environmental variables (Fig. S3). LFMM identified 398 unique loci (~19% of the total loci), with ~30% associated with multiple environmental variables (Fig. S4). Only 22 loci were identified by both LFMM and *Bayenv2*; thus, the final EAA dataset comprised 499 unique loci (Fig. S5).

There were 56 loci present in both the Outlier and EAA datasets (Fig. S6). This represents 45.9% of the Outlier dataset, and 11.2% of the EAA dataset.

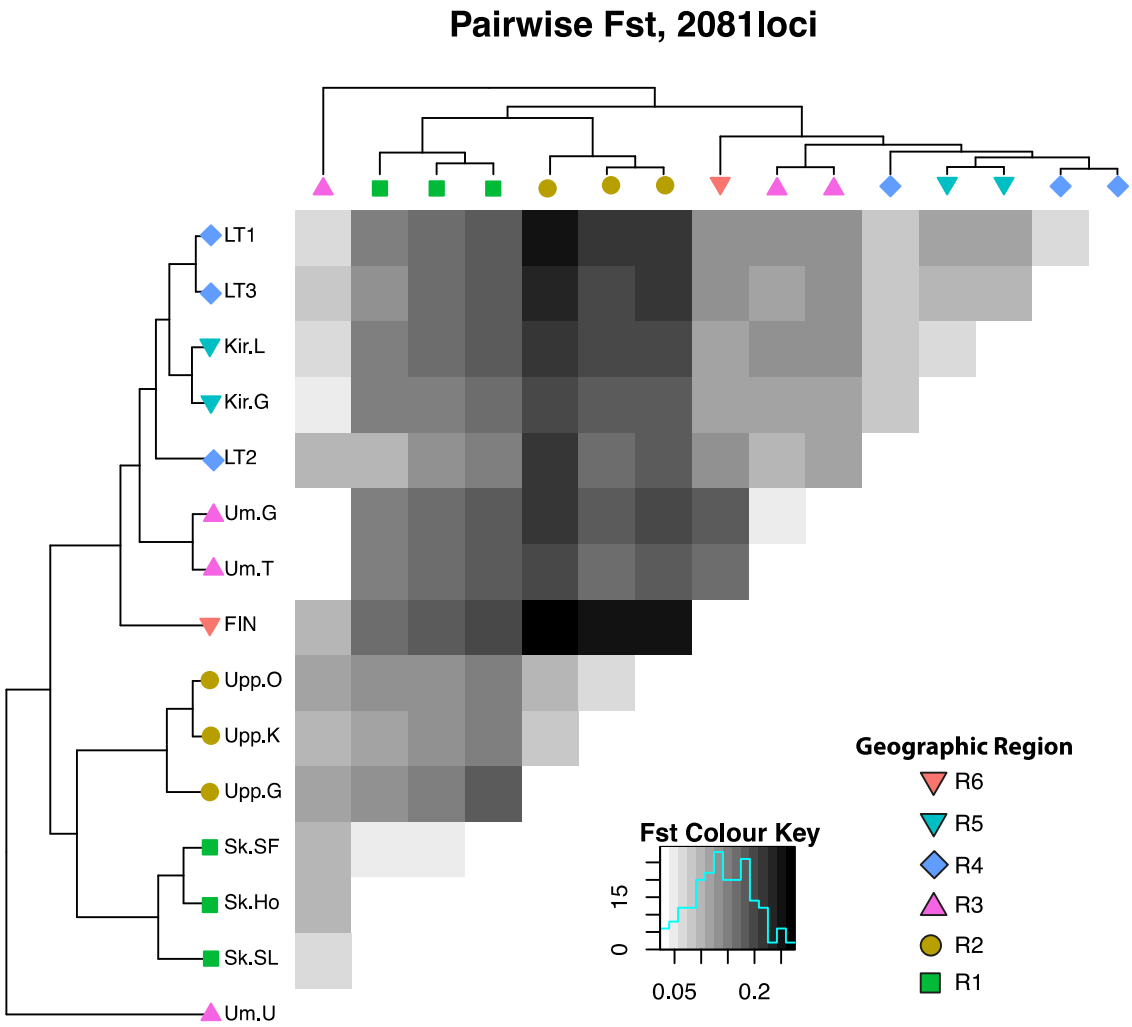


Figure 3 Heatmap of the pairwise genetic divergence (F_{ST}) between all sampled populations. Darker squares indicates higher F_{ST}, with colour scaled as shown in the key. Geographic regions are differentiated with shapes corresponding to populations in Fig. 1. Colours correspond to the PCA shown in Fig. 4. The dendrogram shows the population structure between southern (R1-R2) and northern (R3-R6) Scandinavian populations.

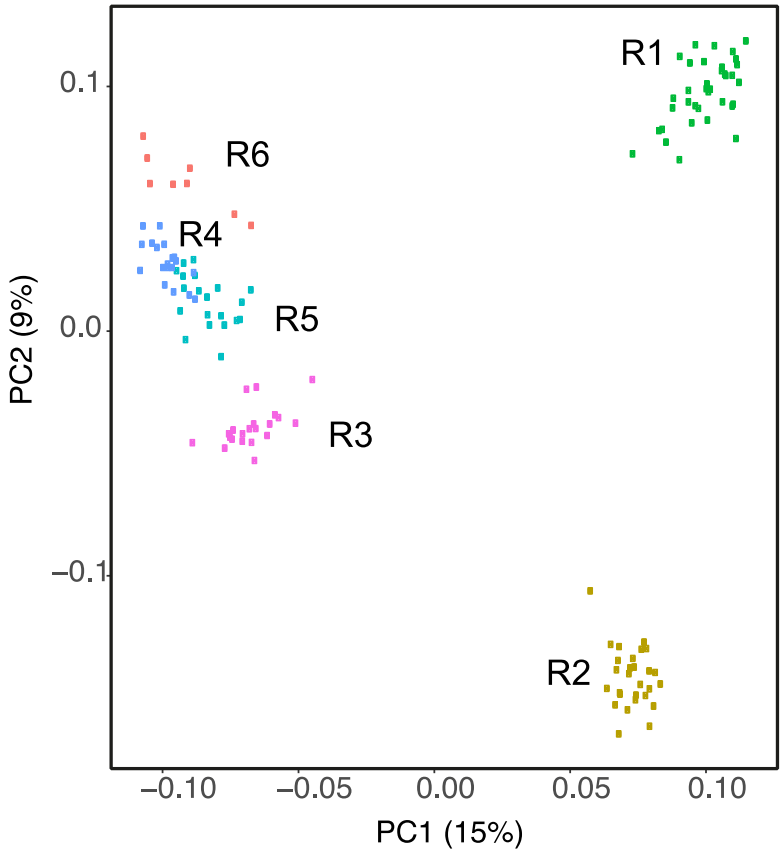


Figure 4 Graph of the first two axes of a Principal Component Analysis of all populations. The proportion variation explained by each axis is shown in brackets. Regions are shown in different colours.

Genomic Turnover - We used the mean R^2 averaged over loci in each dataset as a measure of the fit importance and thus how informative each dataset was (Table 2). The Reference and EAA datasets performed similarly, with mean R^2 values of 37.6% and 38.9%, respectively. The Outlier dataset showed a better fit on average, with a mean R^2 of 56.5%. The frequency distributions of R^2 values differed among datasets. The Reference and EAA datasets both had fairly flat distributions, with 30.5% and 34% of loci with $R^2>0.5$. The R^2 of the Outlier dataset was right skewed, with 67% of loci identified with $R^2>0.5$.

Table 2 Summary of the three SNP datasets and their results from the Gradient Forest analysis. The model was fit only to SNPs that were variable in more than five populations.

SNPs	nr of SNPs	polymorphic in >5 populations	SNPs with R^2 >0.5 (%)	mean % (range)
Reference	1427	1034	292 (30.5%)	37.6 (0.2, 90.0)
Outlier	122	115	77 (67.0%)	56.5 (0.8,91.5)
EAA	499	339	115 (33.9%)	38.9 (0.0,79.0)

SNPs, Single Nucleotide polymorphism

The most important variables for all three datasets were distance and either MEM1 or MEM2 (Fig. 5). This suggests that geographic distance is an important determinant of the genomic differences between populations, and that the MEM variables captured important environmental variation that had not been included in the model. When considering only the BioClim variables, BIO5 and BIO2 were the most important variables for the Outlier dataset, and BIO5 for the EAA dataset. Notably, the Outlier dataset had a higher fit importance (R^2) to the BIO2 and BIO5 variables than the EAA or Reference dataset. This suggests that these variables explain the change in allele frequency in the Outlier dataset better than the change in allele frequency in the Reference our EAA datasets.

We found that three BioClim variables explained a more rapid change in allele frequencies in the adaptive SNP datasets than in the reference datasets. The cumulative importance of allele frequency changes in the Outlier and EAA datasets differed in shape and magnitude for BIO5 (maximum temperature during the warmest month), BIO18 (precipitation during the warmest quarter), and MEM2 (Fig. 6 panels A, D, H). Most notably, the cumulative importance of the Outlier dataset showed big changes in allele frequencies at two points (19°C and 20.5°C) along BIO5 (maximum temperature during the warmest month). A similar, but slightly weaker response was seen in the EAA dataset (arrows in Fig. 6A). To examine this more closely we plotted the change in minor allele frequencies of five of the Outlier loci with the highest relative importance (R^2) associated with BIO5 (Fig. 7A). We find that allele frequencies in R2 and R3 differ dramatically from the rest of the gradient, while

510 frequencies in R1 are similar to R4-R6. This suggests a threshold response to the the
511 higher temperatures experienced by R2 and R3 populations (Fig. 7B), where the
512 adaptive genotype requires a big change in allele frequencies compared to the rest of
513 the gradient. The rate of genomic turnover was less dramatic in response to BIO18
514 (precipitation in the warmest quarter, mm), but Outlier and EAA datasets both
515 showed a higher cumulative importance than the Reference data (arrow in Fig. 6D).
516 Finally, the Outlier dataset showed a higher cumulative importance associated with
517 BIO2 (mean diurnal range in temperature) that was not recovered by the EAA
518 dataset.

519 Spatial mapping of the model for all three datasets showed that genomic
520 turnover was maximal between R1 and R2, followed by the change between R2 and
521 the rest of the populations. The three precipitation variables explained most of the
522 variation in the Reference data allele frequencies (biplots in Fig. 8A, C, and E), while
523 BIO5 (maximum temperature during the warmest month) was important for both the
524 EAA and Outlier datasets. The biggest difference in genomic turnover between the
525 adaptive loci and the Reference dataset was found in southern Sweden (Fig. 9),
526 which suggests that the biggest change in adaptive allele frequencies is required
527 between populations from R1 to R2.

528

529

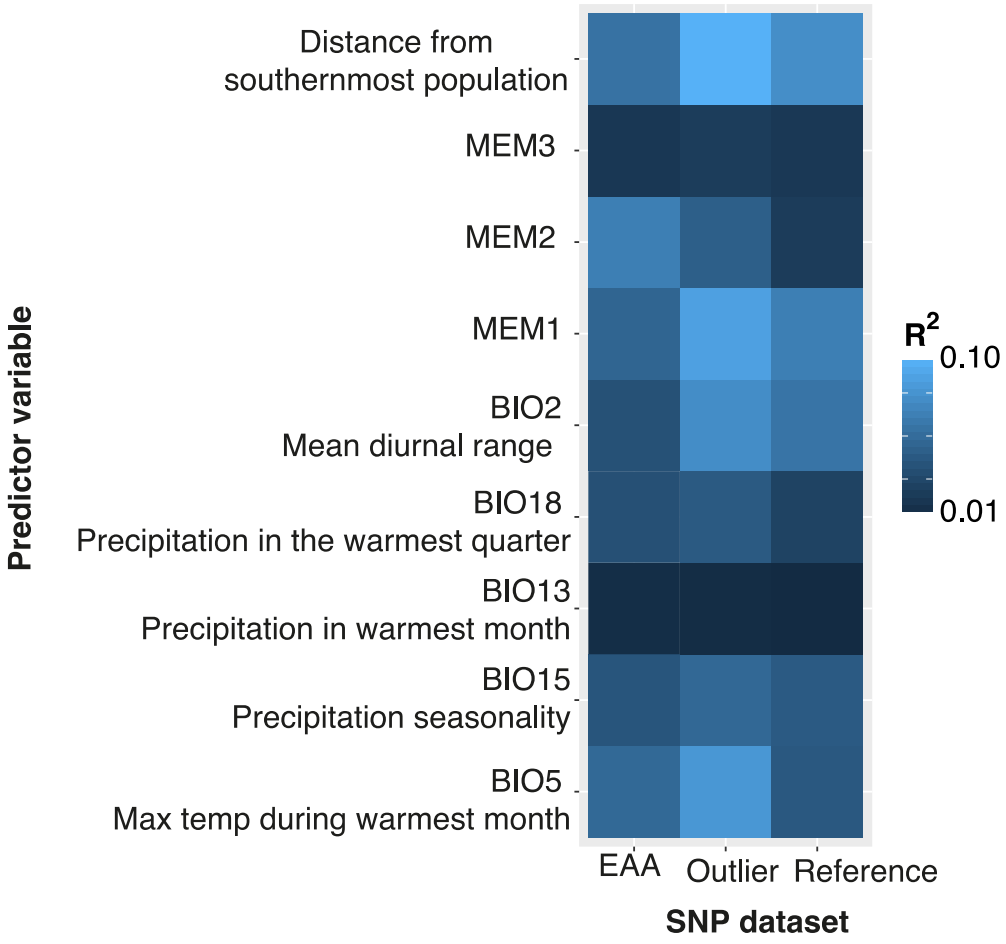


Figure 5 The relative importance (R^2) of the predictor variables used in the Gradient Forest analysis, calculated as the weighted mean proportion of variance in allele frequency explained by a given environmental variable. Results are shown for the five BioClim variables, geographic distance, and three Moran's Eigenvectors (MEM1-3) explaining geographic structure in the dataset. The columns show results for the SNP datasets. EAA: Adaptive loci from environmental association analyses; Outlier: Adaptive loci identified using PCAdapt and X_TX statistic in BayEnv2; Reference: the remaining loci. Lighter colours indicating higher importance; e.g. BIO2 and BIO5 explain a large proportion of variance in the Outlier dataset. The relative importance of variables associated with the Reference dataset provides the null model. Here BIO5 and BIO2 explain more variance in the Outlier SNP dataset than in the Reference dataset, which suggests that these environmental variables are important drivers of adaptive divergence.

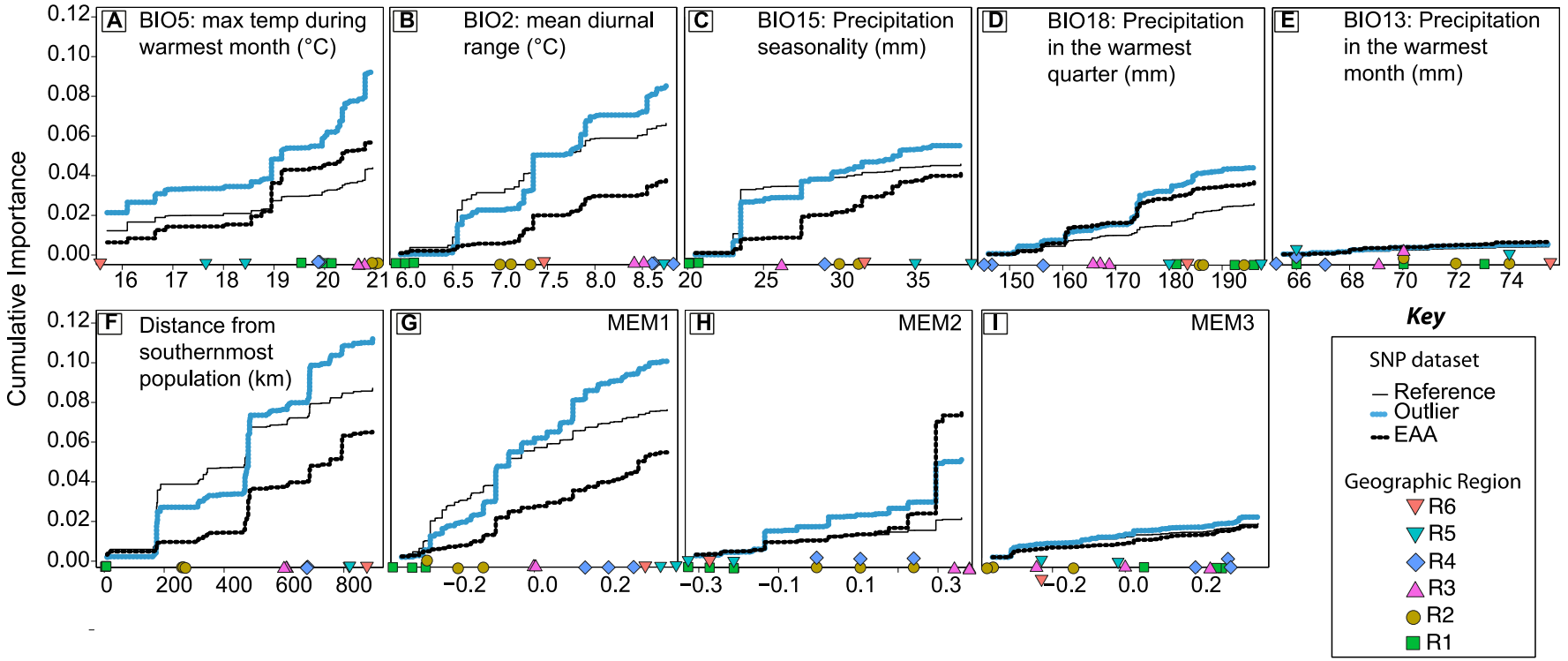
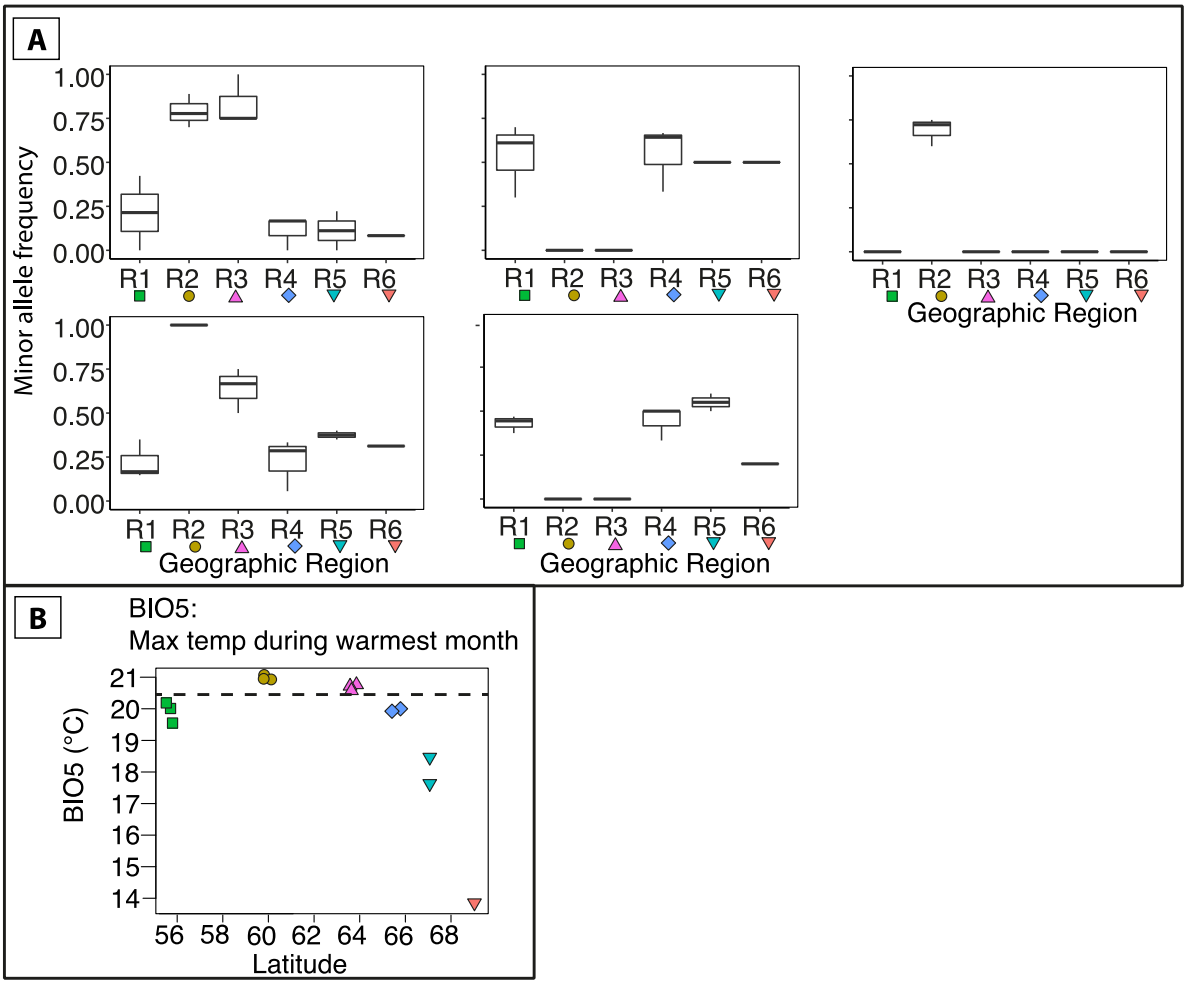


Figure 6 A comparison of the cumulative importance of each predictor variable for the three SNP datasets (see Key). We include the five BioClim variables (A-E) along with geographic distance (F) and three Moran Eigenvector Map variables (G-I) that explain geographic structure in the data. The maximum height of a line indicates the total allelic turnover associated with that variable. The relative importance of a particular point along the predictor variable is seen by the change in line height. The position of all populations along each variable is shown with coloured symbols along the bottom of each graph, with the populations from the same region following the same colour and shape codes as before (see Key). The Reference dataset shows the cumulative importance of each variable in explaining the change in allele frequency of neutral loci.

554



555

556 ***Figure 7*** Standardised change in the minor allele frequency in five loci from the
557 Outlier dataset that were associated with BIO5 is shown across the latitudinal
558 gradient (A). Allele frequencies were all standardised to range between 0 and 1.
559 Allele frequency was found to be dramatically different in R2 and R3 compared to the
560 rest of the gradient. The change in allele frequency between R1 and R2 explains the
561 dramatic difference between the predicted allele frequencies in the Reference and
562 Outlier datasets (Fig. 9). The geographic distribution of BIO5 (B) is shown to illustrate
563 the small change in temperature between R1 and R2. The dashed line indicates the
564 temperature associated with this threshold-like response in the change in allele
565 frequency.

566

567

568

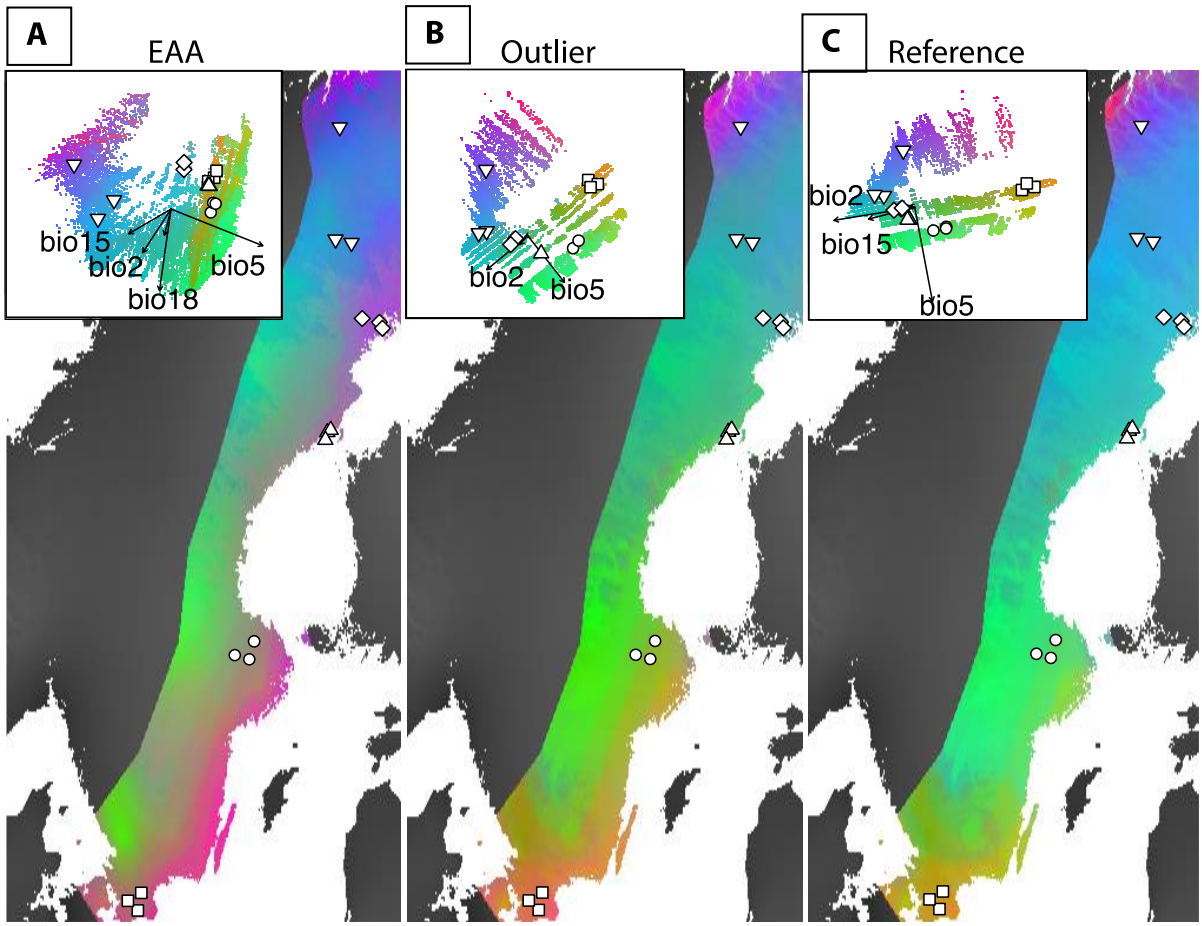


Figure 8 Predicted spatial distribution of genomic composition as determined for the EAA (A), Outlier (B), and Reference (C) datasets. The five BioClim variables are transformed by their relative importance in predicting genomic turnover in each dataset, and visualised as a PCA with a colour assigned to the first three principal components. Population genomic composition is expected to be similar on the same colours. The inset in each panel shows a PCA of the transformed BioClim variables, with the most important variables (see Fig. 5) shown with arrows. Sampled populations are shown on the PCA and map using the same white symbols as in Fig. 1.

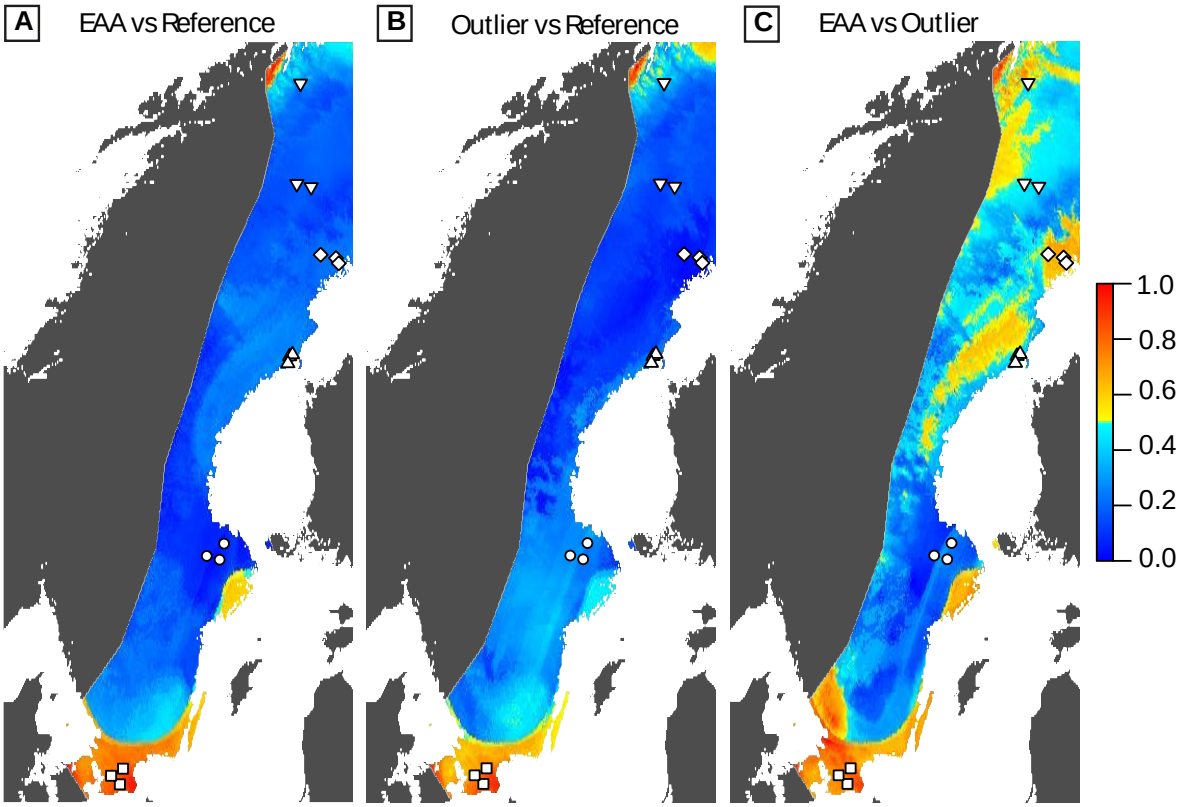


Figure 9 Difference in genomic turnover between the reference and adaptive datasets (a & b), and the two adaptive datasets (c). Distances were calculated as the difference between Procrustes residuals in the matrix comparisons and scaled by the maximum distance found for each comparison. Large differences between datasets are indicated with warmer colours. Sampled populations are shown using white symbols (see Fig. 1).

Discussion

By combining landscape genetics and genomic turnover analyses we describe how climate and geography structure *R. temporaria* genomic variation across the Scandinavian latitudinal gradient. Genomic variation in this system is strongly related to geography, but also shows evidence of adaptation to climate. A surprisingly large portion of the total genomic variation is attributable to climate variables (38%) or geographically structured climate (30%). We also find a threshold response in to BIO5 (maximum temperature in the warmest month), implying that a thermal threshold occurs in southern Sweden. Finally, our results show that the biggest mismatch between neutral and adaptive allele frequencies occurs in southern Sweden, largely driven by the threhold response to BIO5. Our results show that an analysis of the geographic distribution of genomic variation in *R. temporaria* provide important insights into the climatic drivers and potential adaptive thresholds across a well-studies system.

Population Structure

We found strong population structure across the latitudinal gradient, with the biggest divergence in allele frequencies between R1 and the rest of Sweden. Sequential pairing off of the rest of the populations and strong signals of IBD are indicative of an expansion from a southern colonisation. These results support previous work based on mtDNA that has suggested a single colonisation route and northward expansion in Scandinavia by *R. temporaria* (Palo *et al.* 2004). Previous phylogenetic work from the same study indicated that all populations in Scandinavia and northwards assign to the eastern mitochondrial haplogroup (Palo *et al.* 2004), and the contact zone between the eastern and western lineages has been described in Northern Germany (Schmeller *et al.* 2008). However, population assignment analysis based on microsatellite data found that some southern Scandinavian populations (including from R1) assigned up to 100% to the Western lineage based on multilocus genotypes (Palo *et al.* 2004). This explains the divergence of the southern populations, perhaps even up to R2, from the rest of the gradient. Overall our results support previous work that asserted that two mitochondrial lineages

622 colonised Scandinavia from the south; via Denmark and Sweden. A contact zone
623 between these lineages in southern Sweden resulted in the divergent allele
624 frequencies in this region compared to the rest of the gradient. The eastern
625 mitochondrial haplotype occur throughout extant Scandinavian populations, which
626 suggests that gene flow from the populations with the eastern haplogroup have since
627 swamped and replaced the western haplogroup (Palo *et al.* 2004).
628
629 *Geography and Climate determine genotype*
630 We found that climate and geographically structured climate explained a large
631 proportion (76.6%) of the total genomic variation. Climate independent of geography
632 and geographically structured climate explained similar amounts of variation (38%
633 and 30%, respectively). Strong clinal genetic structure across the Scandinavian
634 latitudinal gradient which has been attributed to consistent selection gradients co-
635 varying with geography (Palo *et al.* 2003b, 2004; Cano *et al.* 2004). However, our
636 results suggest that a significant proportion of the adaptive divergence across the
637 gradient could be associated with climate variables that are not latitudinally ordered.
638 Season length (number of days above 6 °C, the developmental threshold for *R.*
639 *temporaria* tadpoles; Laurila *et al.* 2001; Laugen *et al.* 2003b; Muir *et al.* 2014a) and
640 temperature during larval development (30 days after spawning) are two
641 environmental variables that are commonly attributed to the latitudinal adaptive
642 divergence (e.g. Laugen *et al.* 2003b, 2005a). While season length is latitudinally
643 ordered, water temperatures during the larval phase peak at mid-latitudes (Laugen *et*
644 *al.* 2003b). Common garden experiments have found that several larval traits - egg
645 development time, size at hatching, larval growth rate, size at metamorphosis, and
646 resting metabolic rate - follow this curvilinear distribution across latitude (Pahkala *et*
647 *al.* 2002; Laugen *et al.* 2003b, 2005b; Palo *et al.* 2003b; Lindgren & Laurila 2005,
648 2009). Adult body size, skeletal growth, and lifetime activity follow the same
649 curvilinear distribution, and are maximised at mid-latitudes (Laugen *et al.* 2005b;
650 Hjernquist *et al.* 2012).
651 Together with our results show support for latitudinally ordered adaptive divergence,
652 but also present evidence that climatic drivers that are not latitudinally ordered are
653 important for adaptation.

Non-linear changes in allele frequencies and threshold effects

We find a strong threshold effect in a subset of Outlier loci associated with BIO5 (maximum temperature during the warmest month). This suggests that there is a physiological threshold in response to BIO5 (or something related to BIO5). Thesholds in polygenic traits are likely to be common in heterogenous environments (Roff 1996). Indeed, we find that four of the five BioClim variables showed an elevated cumulative importance for the adaptive loci compared with the Reference datasets. Of these, BIO2 (mean diurnal range in temperature) and BIO18 (precipitation during the warmest quarter) show evidence of thresholds at which allele frequencies change more rapidly than in the Reference dataset. Geographically the BIO5 threshold separates R2 and R3 from the rest of the gradient. We found dramatically different allele frequencies in a set of Outlier loci associated with BIO5 in this region. Comparison in genomic turnover between the datasets identified the transition between R1 and R2 to diverge the most between the Reference and adaptive datasets. This is indicative of strong adaptive divergence in this region.

Adaptive divergence across the latitudinal gradient in Europe has been extensively documented in plants and animals (e.g. Laugen *et al.* 2003b; Debieu *et al.* 2013; Vergeer & Kunin 2013). Much of this work is based on common garden and reciprocal transplant experiments have established divergence in various phenotypes across the environmental gradient. Landscape and population genomic methods provide a powerful approach to complement and extend these results in several ways. These include determining the proportion of genomic variation associated with adaptation, identifying the genomic underpinnings and architecture of adaptation, identifying important climatic drivers of adaptive divergence, and identifying adaptive thresholds in response to a specific variable. More generally, this approach can have valuable conservation implications, particularly for mitigating the loss of biodiversity due to climate change. One of the most valuable outcomes lies in identifying populations where a small change in environment will result in a large mismatch between genotype and environment. These populations are particularly vulnerable to extinction, and conservation management action would have to be carefully considered.

686

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR*, pp. 267–281. Budapest:Akadémiai Kiadó.

Alberto F, Niort J, Derory J *et al.* (2010) Population differentiation of sessile oak at the altitudinal front of migration in the French Pyrenees. *Molecular Ecology*, **19**, 2626–2639.

Alton LA, Franklin CE (2017) Drivers of amphibian declines: effects of ultraviolet radiation and interactions with other environmental factors. *Climate Change Responses*, **4**, DOI 10.1186/s40665-017-0034-7.

Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **205**, 403–410.

Alvarez N, Thiel-Egenter C, Tribsch A *et al.* (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632–640.

Ammann B, Birks HJB, Brooks SJ *et al.* (2000) Quantification of biotic responses to rapid climatic changes around the Younger Dryas — a synthesis. *Paleogeography, Paleoclimatology, Paleoecology*, **159**, 313–347.

Avise JC, Riddle B (2009) Phylogeography : Retrospect and Prospect. *Journal of Biogeography*, **36**, 3–15.

Bachmann JC (2017) Adaptive Divergence across an Elevational Gradient in the Common Frog (Rana temporaria). University of Zurich.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Molecular ecology*, **17**, 3583–4.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (Rana temporaria). *Molecular biology and evolution*, **23**, 773–83.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Braaker S, Heckel G (2009) Transalpine colonisation and partial

phylogeographic erosion by dispersal in the common vole (Microtus arvalis). *Molecular Ecology*, **18**, 2528–2531.

Brady LD, Griffiths RA (2000) Developmental responses to pond desiccation in tadpoles of the British anuran amphibians (Bufo bufo, B. calamita and Rana temporaria). *Journal of Zoology, London*, **252**, 61–69.

Van Buskirk J (2012) Permeability of the landscape matrix between amphibian breeding sites. *Ecology and evolution*, **2**, 3160–7.

Canestrelli D, Cimmaruta R, Costantini V, Nascetti G (2006) Genetic diversity and phylogeography of the Apennine yellow-bellied toad Bombina pachypus, with implications for conservation. *Molecular Ecology*, **15**, 3741–3754.

Canestrelli D, Cimmaruta R, Nascetti G (2008) Population genetic structure and diversity of the Apennine endemic stream frog, Rana italica – insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Molecular Ecology*, **17**, 3856–3872.

Canestrelli D, Nascetti G (2008) Phylogeography of the pool frog Rana (Pelophylax) lessonae in the Italian peninsula and Sicily: Multiple refugia, glacial expansions and nuclear-mitochondrial discordance. *Journal of Biogeography*, **35**, 1923–1936.

Cano JM, Laurila A, Palo J, Merilä J (2004) Population differentiation in G matrix structure due to natural selection in Rana temporaria. *Evolution*, **58**, 2013–2020.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3:Genes, Genomes, Genetics*, **1**, 171–182.

Caye K, Deist TM, Martins H, Michel O, François O (2016) TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, **16**, 540–548.

CH2014–Impacts (2014) *Toward quantitative scenarios of climate change Impacts in Switzerland*. OCCR, FOEN, MeteoSwiss, C2SM, Agroscope, and ProClim, Bern, Switzerland.

Clement M, Posada D, Crandall KA (2000) TCS : a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Cornetti L, Lemoine M, Hilfiker D *et al.* (2016) Higher genetic diversity on mountain tops: the role of historical and contemporary processes in shaping genetic variation in the bank vole. *Biological Journal of the Linnean Society*, **118**, 233–244.

760 Croteau MC, Davidson MA, Lean DRS, Trudeau VL (2008) Global Increases in
761 Ultraviolet B Radiation : Potential Impacts on Amphibian Development and
762 Metamorphosis. *Physiological and Biochemical Zoology*, **81**, 743–761.

763 Crottini A, Andreone F, Kosuch J *et al.* (2007) Fossorial but widespread: the
764 phylogeography of the common spadefoot toad (Pelobates fuscus), and the
765 role of the Po Valley as a major source of genetic variability. *Molecular*
766 *Ecology*, **16**, 2734–2754.

767 Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and
768 VCFtools. *Bioinformatics*, **27**, 2156–2158.

769 Dansgaard W, Johnsen SJ, Clausen HB *et al.* (1993) Evidence for general
770 instability of past climate from a 250-kyr ice-core record. *Nature*, **364**, 218–
771 220.

772 Darnault R, Rolland Y, Braucher R *et al.* (2011) Timing of the last deglaciation
773 revealed by receding glaciers at the Alpine-scale : impact on mountain
774 geomorphology. *Quaternary Science Reviews*, 1–16.

775 Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more
776 models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

777 Debieu M, Tang C, Stich B *et al.* (2013) Co-Variation between seed dormancy,
778 growth rate and flowering time changes with latitude in Arabidopsis
779 thaliana. *PloS one*, **8**, 1–12.

780 Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA Phylogeography of
781 the Common Beech (Fagus sylvatica L.) in Europe. *Evolution*, **50**, 2515–
782 2520.

783 Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics
784 with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**,
785 1969–1973.

786 Dudaniec RY, Spear SF, Richardson JS, Storfer A (2012) Current and historical
787 drivers of landscape genetic structure differ in core and peripheral
788 salamander populations. *PloS one*, **7**.

789 Eaton DAR (2014) PyRAD: Assembly of de novo RADseq loci for phylogenetic
790 analyses. *Bioinformatics*, **30**, 1844–1849.

791 Ellis N, Smith SJ, Pitcher CR (2012) Gradient forests : calculating importance
792 gradients on physical predictors. *Ecology*, **93**, 156–168.

793 Excoffier L, Lischer HEL (2010) An Integrated Software Package for Population
794 Genetics Data Analysis. *Molecular Ecology Resources*, **10**, 564–567.

795 Fitzpatrick SW, Gerberich JC, Kronenberger JA, Angeloni LM, Funk WC (2015)

796 Locally adapted traits maintained in the face of high gene flow. *Ecology*
797 *Letters*, **18**, 37–47.

798 Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level
799 modelling of biodiversity: mapping the genomic landscape of current and
800 future environmental adaptation. *Ecology Letters*, **18**, 1–16.

801 Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting
802 spatial genetic signatures of local adaptation in heterogeneous landscapes.
803 *Molecular Ecology*, **25**, 104–120.

804 Fox J, Weisberg S (2011) *An {R} Companion to Applied Regression*. Thousand
805 Oaks CA: Sage.

806 Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic
807 signals of diversification along ecological gradients in a tropical lizard.
808 *Molecular ecology*, **19**, 3773–3788.

809 Frichot E, François O (2015) LEA : An R package for landscape and ecological
810 association studies (B O'Meara, Ed,). *Methods in Ecology and Evolution*, **6**,
811 925–929.

812 Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for
813 associations between loci and environmental gradients using latent factor
814 mixed models. *Molecular Biology and Evolution*, **30**, 1687–99.

815 Frichot E, Schoville S, de Villemereuil P, Gaggiotti OE, François O (2015)
816 Detecting adaptive evolution based on association with ecological
817 gradients: Orientation matters! *Heredity*, **115**, 22–28.

818 Garcia VOS, Ivy C, Fu J (2017) Syntopic frogs reveal different patterns of
819 interaction with the landscape: A comparative landscape genetic study of
820 Pelophylax nigromaculatus and Fejervarya limnocharis from central China.
821 *Ecology and Evolution*, **7**, 9294–9306.

822 Gosner KL (1960) A Simplified Table for Staging Anuran Embryos and Larvae
823 with Notes on Identification. *Herpetologica*, **16**, 183–190.

824 Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-
825 statistics. *Molecular Ecology Notes*, **5**, 184–186.

826 Gugerli F, Englisch T, Niklfeld H *et al.* (2008) Relationships among levels of
827 biodiversity and the relevance of intraspecific diversity in conservation – a
828 project synopsis. *Perspectives in Plant Ecology, Evolution and*
829 *Systematics*, **10**, 259–281.

830 Günther T, Coop G (2013) Robust identification of local adaptation from allele
831 frequencies. *Genetics*, **195**, 205–220.

832 Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and
833 analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*,
834 **41**, 95–98.

835 Harrisson KA, Amish SJ, Pavlova A *et al.* (2017) Signatures of polygenic
836 adaptation associated with climate across the range of a threatened fish
837 species with high genetic connectivity. *Molecular ecology*, **26**, 6253–6269.

838 Hecht BC, Matala AP, Hess JE, Narum SR (2015) Environmental adaptation in
839 Chinook salmon (Oncorhynchus tshawytscha) throughout their North
840 American range. *Molecular Ecology*, **24**, 5573–5595.

841 Heiri C, Bugmann H, Tinner W, Heiri O, Lischke H (2006) A model-based
842 reconstruction of Holocene treeline dynamics in the Central Swiss Alps.
843 *Journal of Ecology*, **94**, 206–216.

844 Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics
845 of adaptation from standing genetic variation. *Genetics*, **169**, 2335–52.

846 Hewitt GM (1996) Some genetic consequences of ice ages, and their role in
847 divergence and speciation. *Biological Journal of the Linnean Society*, **58**,
848 247–276.

849 Hewitt G (1999) Post-glacial re-colonization of European biota. *Biological*
850 *Journal of the Linnean Society*, **68**, 87–112.

851 Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**,
852 907–913.

853 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high
854 resolution interpolated climate surfaces for global land areas. *International*
855 *Journal of Climatology*, **25**, 1965–1978.

856 Hitchings SP, Beebee TJC (1997) Genetic substructuring as a result of barriers
857 to gene flow in urban Rana temporaria (common frog) populations:
858 implications for biodiversity conservation. *Heredity*, **79**, 117–127.

859 Hjernquist MB, Soderman F, Jonsson KI *et al.* (2012) Seasonality determines
860 patterns of growth and age structure over a geographic gradient in an
861 ectothermic vertebrate. *Oecologia*, **170**, 641–649.

862 Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the Genomic Basis of
863 Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The*
864 *American Naturalist*, **188**, 000–000.

865 Hoffmann A a, Sgrò CM (2011) Climate change and evolutionary adaptation.
866 *Nature*, **470**, 479–85.

867 Huelsenbeck JP, Ronquist F (2001) MRBAYES : Bayesian inference of

phylogenetic trees. *Bioinformatics Application Note*, **17**, 754–755.

Ilut DC, Nydam ML, Hare MP (2014) Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and diagnostics for optimal clustering. *BioMed Research International*, **2014**, 9 pages.

Jombart T (2008) Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Jones MR, Forester BR, Teufel AI *et al.* (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution*, **67**, 3455–68.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Lasky JR, Des Marais DL, McKay J *et al.* (2012) Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. *Molecular Ecology*, **21**, 5512–5529.

Laugen AT, Kruuk LEB, Laurila A *et al.* (2005a) Quantitative genetics of larval life-history traits in Rana temporaria in different environmental conditions. *Genetics Research*, **86**, 161–170.

Laugen AT, Laurila A, Jönsson KI, Söderman F, Merilä J (2005b) Do common frogs (*Rana temporaria*) follow Bergmann's rule? *Evolutionary Ecology Research*, **7**, 717–731.

Laugen AT, Laurila A, Merilä J (2002) Maternal and genetic contributions to geographical variation in Rana temporaria larval life-history traits. *Biological Journal fo the Linnean Society*, **76**, 61–70.

Laugen AT, Laurila A, Merilä J (2003a) Latitudinal and temperature-dependent variation in embryonic development and growth in Rana temporaria. *Oecologia*, **135**, 548–554.

Laugen AT, Laurila A, Rasanen K, Merilä J (2003b) Latitudinal countergradient variation in the common frog (Rana temporaria) development rates – evidence for local adaptation. *Journal of Evolutionary Biology*, **16**, 996–1005.

Laurila A, Karttunen S, Meril J (2002) Adaptive Phenotypic Plasticity and Genetics of Larval Life Histories in Two Rana Temporaria Populations. *Evolution*, **56**, 617–627.

Laurila A, Pakkasmaa S, Merila J (2001) Influence of Seasonal Time Constraints on Growth and Development of Common Frog Tadpoles: A Photoperiod Experiment. *Oikos*, **95**, 451–460.

Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.

Legendre P, Legendre LF (2012) *Numerical ecology*. Elsevier.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics Application Note*, **25**, 1451–1452.

Lindgren B, Laurila A (2005) Proximate causes of adaptive growth rates: growth efficiency variation among latitudinal populations of Rana temporaria. *Journal of Evolutionary Biology*, **18**, 820–828.

Lindgren B, Laurila A (2009) Physiological variation along a geographical gradient: is growth rate correlated with routine metabolic rate in Rana temporaria tadpoles? *Biological Journal of the Linnean Society*, **98**, 217–224.

Loman J, Claesson D (2003) Plastic response to pond drying in tadpoles Rana temporaria: tests of cost models. *Evolutionary Ecology Research*, **5**, 179–194.

Lugon-Moulin N, Hausser J (2002) Phylogeographical structure, postglacial recolonization and barriers to gene flow in the distinctive Valais chromosome race of the common shrew (Sorex araneus). *Molecular Ecology*, **11**, 785–794.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–94.

Luu K, Bazin E, Blum MGB (2016) pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, **33**, 67–77.

Manel S, Perrier C, Pratlong M *et al.* (2016) Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular ecology*, **25**, 170–184.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. *Molecular ecology*, **19**, 3824–35.

Marquis O, Miaud C (2008) Variation in UV sensitivity among common frog Rana temporaria populations along an altitudinal gradient. *Zoology*, **111**,

309–317.

Marquis O, Miaud C, Lena J-P (2008) Developmental responses to UV-B radiation in common frog Rana temporaria embryos from along an altitudinal gradient. *Population Ecology*, **50**, 123–130.

Mátyás G, Sperisen C (2001) Chloroplast DNA polymorphisms provide evidence for postglacial re-colonisation of oaks (Quercus spp.) across the Swiss Alps. *Theoretical and Applied Genetics*, **102**, 12–20.

McCain CM, Colwell RK (2011) Assessing the threat to montane biodiversity from discordant shifts in temperature and precipitation in a changing climate. *Ecology Letters*, **14**, 1236–1245.

Merila J, Laurila A, Laugen AT, Rasanen K, Pahkala M (2000) Plasticity in age and size at metamorphosis in Rana temporaria: comparison of high and low latitude populations. *Ecography*, **23**, 457–465.

Merilä J, Laurila A, Laugen AT, Räsänen K, Pahkala M (2000) Plasticity in Age and Size at Metamorphosis in Rana Temporaria: Comparison of High and Low Latitude Populations. *Ecography*, **23**, 457–465.

Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution*, **28**, 659–669.

Miaud C, Guyetant R, Elmberg J (1999) Variations in life-history traits in the common frog Rana temporaria (Amphibia: Anura): a literature review and new data from the French Alps. *Journal of Zoology*, **249**, 61–73.

Miaud C, Merilä J (2001) Local adaptation or environmental induction? Causes of population differentiation in alpine amphibians. *Biota*, **2**, 31–50.

Muir AP, Biek R, Mable BK (2014a) Behavioural and physiological adaptations to low-temperature environments in the common frog, Rana temporaria. *BMC Evolutionary Biology*, **14**, 1471–2148.

Muir A, Piek R, Thomas R, Mable B (2014b) Local adaptation with high gene flow: temperature parameters drive adaptation to altitude in the common frog (Rana temporaria). *Molecular ecology*, **23**, 561–574.

Nychka D, Furrer R, Paige J, Sain S (2015) fields: Tools for spatial data.

Oksanen J, Blanchet FG, Kindt R *et al.* (2015) Vegan: community ecology package. R package vegan, vers. 2.2-1.

Pahkala M, Laurila A, Meril J (2002) Effects of ultraviolet-B radiation on common frog Rana temporaria embryos from along a latitudinal gradient. *Oecologia*, **133**, 458–465.

Pahkala M, Laurila A, Merila J (2000) Ambient Ultraviolet-B radiation reduces

977 hatchling size in the common frog Rana temporaria. *Ecography*, **23**, 531–
978 538.

979 Pahkala M, Laurila A, Merila J (2001) Carry-over effects of ultraviolet-B
980 radiation on larval tness in Rana temporaria. *Proceedings of the Royal*
981 *Society of London B: Biological Sciences*, **268**, 1699–1706.

982 Pahkala M, Merila J, Ots I, Laurila A (2003) Effects of ultraviolet-B radiation on
983 metamorphic traits in the common frog Rana temporaria. *Journal of the*
984 *Zoological Society of London*, **259**, 57–62.

985 Palo JU, O'Hara RB, Laugen AT *et al.* (2003a) Latitudinal divergence of
986 common frog (Rana temporaria) life history traits by natural selection:
987 evidence from a comparison of molecular and quantitative genetic data.
988 *Molecular ecology*, **12**, 1963–1978.

989 Palo JU, O'Hara RB, Laugen AT *et al.* (2003b) Latitudinal divergence of the
990 common frog (Rana temporaria) life history traits by natural selection:
991 evidence from a comparison of molecular quantitative genetic data.
992 *Molecular Ecology*, **12**, 1963–1978.

993 Palo JU, Schmeller DS, Laurila A *et al.* (2004) High degree of population
994 subdivision in a widespread amphibian. *Molecular Ecology*, **13**, 2631–2644.

995 Parisod C (2008) Postglacial recolonisation of plants in the western Alps of
996 Switzerland. *Botanica Helvetica*, **118**, 1–12.

997 Parmesan C (2006) Ecological and Evolutionary Responses to Recent Climate
998 Change. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 637–
999 669.

1000 Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change
1001 impacts across natural systems. *Nature*, **421**, 37–42.

1002 Pennings PS, Hermisson J (2006) Soft sweeps II - Molecular population
1003 genetics of adaptation from recurrent mutation or migration. *Molecular*
1004 *Biology and Evolution*, **23**, 1076–1084.

1005 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double
1006 Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and
1007 Genotyping in Model and Non-Model Species. *PloS one*, **7**.

1008 Polechová J, Barton NH (2015) Limits to adaptation along environmental
1009 gradients. *Proceedings of the National Academy of Sciences of the United*
1010 *States of America*, **112**, 6401–6406.

1011 Poncet BN, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological
1012 relevance using a genome scan in two independent regional population
1013 samples of Arabis alpina. *Molecular ecology*, **19**, 2896–907.

1014 Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature*
1015 *Reviews Genetics*, **11**, 665–7.

1016 Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: A tool set for whole-
1017 genome association and population-based linkage analyses. *American*
1018 *Journal of Human Genetics*, **81**, 559–575.

1019 Raj A, Stephens M, Pritchard JK (2014) FastSTRUCTURE: Variational
1020 inference of population structure in large SNP data sets. *Genetics*, **197**,
1021 573–589.

1022 Rehm EM, Olivas P, Stroud J, Feeley KJ (2015) Losing your edge: climate
1023 change and the conservation value of range-edge populations. *Ecology*
1024 *and evolution*, **5**, 4315–4326.

1025 Rellstab C, Fischer MC, Zoller S *et al.* (2016) Local adaptation (mostly) remains
1026 local: reassessing environmental associations of climate-related candidate
1027 SNPs in Arabidopsis halleri. *Heredity*, 1–9.

1028 Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A
1029 practical guide to environmental association analysis in landscape
1030 genomics. *Molecular Ecology*, **24**, 4348–70.

1031 Rigling A, Bigler C, Eilmann B *et al.* (2013) Driving factors of a vegetation shift
1032 from Scots pine to pubescent oak in dry Alpine forests. *Global Change*
1033 *Biology*, **19**, 229–240.

1034 Rodrigues N, Betto-Colliard C, Jourdan-Pineau H, Perrin N (2013) Within-
1035 population polymorphism of sex-determination systems in the common frog
1036 (Rana temporaria). *Journal of evolutionary biology*, **26**, 1569–1577.

1037 Roesti M, Salzburger W, Berner D (2012) Uninformative polymorphisms bias
1038 genome scans for signatures of selection. *BMC Evolutionary Biology*, **12**,
1039 94.

1040 Roff DA (1996) The evolution of threshold traits in animals. *The Quarterly*
1041 *Review of Biology*, **71**, 3–35.

1042 Rogers AR, Harpending H (1992) Population Growth Makes Waves in the
1043 Distribution of Pairwise Genetic Differences. *Molecular biology and*
1044 *evolution*, **9**, 552–569.

1045 Rogivue A, Graf R, Parisod C, Holderegger R, Gugerli F (2018) The
1046 phylogeographic structure of Arabis alpina in the Alps shows consistent
1047 patterns across different types of molecular markers and geographic
1048 scales. *Alpine Botany*, **0**, 0.

1049 Rousset F (1997) Genetic differentiation and estimation of gene flow from F-
1050 statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

1051 Roy K, Valentine JW, Jablonski D, Kidwell SM (1996) Scales of climatic
1052 variability and time averaging in Pleistocene biotas: implications for ecology
1053 and evolution. *Trends in Ecology & Evolution*, **11**, 458–463.

1054 Schmeller DS, Palo JU, Merilä J (2008) A contact zone between two distinct
1055 Rana temporaria lineages in northern Germany. *Alytes*, **25**, 93–98.

1056 Schneider S, Excoffier L (1999) Estimation of Past Demographic Parameters
1057 From the Distribution of Pairwise Differences When the Mutation Rates
1058 Vary Among Sites: Application to Human Mitochondrial DNA. *Genetics*,
1059 **152**, 1079–1089.

1060 Schweizer RM, VonHoldt BM, Harrigan R *et al.* (2016) Genetic subdivision and
1061 candidate genes under selection in North American grey wolves. *Molecular*
1062 *Ecology*, **25**, 380–402.

1063 Sillero N, Campos J, Bonardi A *et al.* (2014) Updated distribution and
1064 biogeography of amphibians and reptiles of Europe. *Amphibia-Reptilia*, **35**,
1065 1–31.

1066 Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the
1067 genomics of trees: approaches for understanding local adaptation and
1068 population responses to changing climate. *Tree Genetics & Genomes*, **9**,
1069 901–911.

1070 Ståhlberg F, Olsson M, Uller T (2001) Population divergence of developmental
1071 thermal optima in Swedish common frogs, Rana temporaria. *Journal of*
1072 *Evolutionary Biology*, **14**, 755–762.

1073 Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic
1074 analyses with thousands of taxa and mixed models. *Bioinformatics*
1075 *Application Note*, **22**, 2688–2690.

1076 Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next
1077 generation. *Trends in ecology & evolution*, **25**, 705–12.

1078 Stefani F, Gentilli A, Sacchi R *et al.* (2012) Refugia within refugia as a key to
1079 disentangle the genetic pattern of a highly variable species: The case of
1080 Rana temporaria Linnaeus, 1758 (Anura, Ranidae). *Molecular*
1081 *Phylogenetics and Evolution*, **65**, 718–726.

1082 Stehlik I, Blattner FR, Holderegger R, Bachmann K (2002) Nunatak survival of
1083 the high Alpine plant Eritrichium nanum (L.) Gaudin in the central Alps
1084 during the ice ages. *Molecular Ecology*, **11**, 2027–2036.

1085 Taberlet P, Fumagalli L, Wust-Saucy A-G, Cossons J-F (1998) Comparative
1086 phylogeography and postglacial colonization routes in Europe. *Molecular*
1087 *Ecolgy*, **7**, 453–464.

1088 Teacher AGF, Garner TWJ, Nichols RA (2009) European phylogeography of
1089 the common frog (Rana temporaria): routes of postglacial colonization into
1090 the British Isles, and evidence for an Irish glacial refugium. *Heredity*, **102**,
1091 490–6.

1092 Templeton AR, Crandall KA, Sing CF (1992) Cladistic Analysis of Phenotypic
1093 Associations With Haplotypes Inferred From Restriction Endonuclease
1094 Mapping and DNA Sequence Data. III. Cladogram Estimation. *Genetics*,
1095 **132**, 619–633.

1096 Thomassen H a, Cheviron Z a, Freedman AH *et al.* (2010) Spatial modelling
1097 and landscape-level approaches for visualizing intra-specific variation.
1098 *Molecular Ecology*, **19**, 3532–48.

1099 Tinner W, Theurillat J (2003) Uppermost limit, extent, and fluctuations of the
1100 timberline and treeline ecocline in the Swiss Central Alps during the past
1101 11,500 years. *Arctic, Antarctic, and Alpine Research*, **35**, 158–169.

1102 Veith M, Kosuch J, Vences M (2003) Climatic oscillations triggered post-
1103 Messinian speciation of Western Palearctic brown frogs (Amphibia,
1104 Ranidae). *Molecular phylogenetics and evolution*, **26**, 310–327.

1105 Veith M, Vences M, Vieites DR, Nieto-roman S, Palanca A (2002) Genetic
1106 differentiation and population structure within Spanish common frogs (Rana
1107 temporaria complex; Ranidae, Amphibia). *Folia Zoologica*, **51**, 307–318.

1108 Vences M, Hauswaldt JS, Steinfartz S *et al.* (2013) Radically different
1109 phylogeographies and patterns of genetic variation in two European brown
1110 frogs, genus Rana. *Molecular phylogenetics and evolution*, **68**, 657–70.

1111 Vergeer P, Kunin WE (2013) Adaptation at range margins: common garden
1112 trials and the performance of Arabidopsis lyrata across its northwestern
1113 European range. *New Phytologist*, **197**, 989–1001.

1114 Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting Natural Selection in
1115 Genomic Data. *Annu. Rev. Genet*, **47**, 97–120.

1116 Wang IJ (2012) Environmental and topographic variables shape genetic
1117 structure and effective population sizes in the endagered Yosemite toad.
1118 *Diversity and Distributions*, **18**, 1033–1041.

1119 Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and
1120 geography in spatial genetic divergence. *Ecology letters*, **16**, 175–82.

1121 Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of
1122 Population Structure. *Evolution*, **38**, 1358–1370.

1123 Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-
1124 Weinberg equilibrium. *American Journal of Human Genetics*, **76**, 887–93.

1125 Willis KJ, Whittaker RJ (2008) The refugial debate. *Science*, **287**, 1406–1407.

1126 Yannic G, Basset P, Hausser J (2008) Phylogeography and recolonization of
1127 the Swiss Alps by the Valais shrew (Sorex antinorii), inferred with
1128 autosomal and sex-specific markers. *Molecular Ecology*, **17**, 4118–4133.

1129 Yannic G, Pellissier L, Dubey S *et al.* (2012) Multiple refugia and barriers
1130 explain the phylogeography of the Valais shrew, Sorex antinorii (Mammalia:
1131 Soricomorpha). *Biological Journal of the Linnean Society*, **105**, 864–880.

1132 Yeaman S (2015) Local Adaptation by Alleles of Small Effect. *The American*
1133 *Naturalist*, **186**, S74–S89.

1134