MR. PATRIK RÖDIN-MÖRCH (Orcid ID : 0000-0001-6737-1488)

PROF. JACOB HÖGLUND (Orcid ID : 0000-0002-5840-779X)

DR. A LAURILA (Orcid ID : 0000-0001-8090-3776)

Article type : Original Article

Genetic basis of amphibian larval development along a latitudinal gradient: gene diversity, selection and links with phenotypic variation in transcription factor C/EBP-1

Yvonne Meyer-Lucht¹, Emilien Luquet², Fríða Jóhannesdóttir^{3,4}, Patrik Rödin-Mörch¹, María Quintela⁵, Alex Richter-Boix¹, Jacob Höglund¹, Anssi Laurila^{1*}

¹ Animal Ecology/Department of Ecology and Genetics, Uppsala University, Uppsala 75236, Sweden

² Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, Université Lyon 1, Villeurbanne Cedex 69622, France

³ Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853, USA

⁴ Ecology and Genetics Research Unit, University of Oulu, Oulu 90014, Finland.

⁵ Department of Population Genetics, Institute of Marine Research, Bergen 5817, Norway

*Author for correspondence: anssi.laurila@ebc.uu.se

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.15123

This article is protected by copyright. All rights reserved.

Abstract

Ectotherm development rates often show adaptive divergence along climatic gradients, but the genetic basis for this variation is rarely studied. Here, we investigated the genetic basis for phenotypic variation in larval development in the moor frog *Rana arvalis* from five regions along a latitudinal gradient from Germany to northern Sweden. We focused on the C/EBP-1 gene, a transcription factor associated with larval development time. Allele frequencies at C/EBP-1 varied strongly among geographic regions. Overall, the distribution of alleles along the gradient was in concordance with the dual postglacial colonization routes into Scandinavia, with a large number of alleles exclusively present along the southern colonization route. Only three of 38 alleles were shared between the routes. Analysis of contemporary selection on C/EBP-1 showed divergent selection among the regions, likely reflecting adaptation to the local environmental conditions, although this was especially strong between southern and northern regions coinciding also with lineages from different colonization routes. Overall, the C/EBP-1 gene has historically been under purifying selection, however, two specific amino acid positions showed significant signals of positive selection. These positions showed divergence between southern and northern regions, and we suggest that they are functionally involved in the climatic adaptation of larval development. Using phenotypic data from a common garden experiment, we found evidence for specific C/EBP-1 alleles being correlated with larval development time, suggesting a functional role in adaptation of larval development to large-scale climatic variation.

Introduction

Understanding the underlying genetic mechanisms for climatic adaptation and assessing the adaptive potential of populations has become an urgent necessity with global climate change (Savolainen *et al.* 2013). Ultimately, insights into the genetic basis of climatic adaptation may help us to predict how populations will be able to adapt to climate change and to identify the potentially endangered

populations (Savolainen *et al.* 2013; Wilczek *et al.* 2009). Here, environmental gradients provide a powerful tool in ecological and evolutionary research (Endler 1977). In studies of climatic adaptation, comparisons of populations distributed along latitudinal or altitudinal gradients are a powerful, widely used method demonstrating that natural selection stemming from climatic variation is a strong selective force, which often leads to local adaptation (e.g. Angilletta 2009; Conover *et al.* 2009; Kawecki & Ebert 2004; Keller *et al.* 2013). However, while studies identifying the genes behind adaptive divergence in wild populations are accumulating (e.g. Hoekstra *et al.* 2006; Jones *et al.* 2012; Lamichhaney *et al.* 2012; Poelstra *et al.* 2014), the genetic basis of the vast majority of traits remains unknown. This is the case also for two major fitness traits, growth and development rates, in which adaptive latitudinal and altitudinal variation among natural populations is common (Conover *et al.* 2009; Dmitriew 2011), but where we, apart from a few studies (e.g., Bonin *et al.* 2005; Guo *et al.* 2016; Page *et al.* 2013; Voss *et al.* 2012; Yang *et al.* 2012), lack insights on the genetic background of the phenotypic variation.

Although much can be learned about adaptive processes without knowledge of specific loci, identifying the genetic features underlying local adaptation is, in many instances, the critical step in addressing fundamental questions in evolutionary biology (Hoban *et al.* 2016; Rausher & Delph 2015). Towards this goal, genetic differentiation outlier methods and selection analyses are powerful and widely used tools for candidate gene detection (Hoban *et al.* 2016). As a complementary approach, association analyses are commonly used, linking phenotypic and genotypic information by testing if a certain genotype or haplotype is more frequently associated with a phenotypic trait than by chance alone (Vasemägi & Primmer 2005). Given there is solid evidence of the gene function, candidate gene association analyses are recognized as a valuable means of identifying the molecular basis controlling phenotypic traits (Vasemägi & Primmer 2005). However, this type of analysis usually identifies large effect loci while missing loci with small effects on the phenotypic trait (Lee *et al.* 2014). Once candidate genes are identified, the ultimate goal is to reveal the molecular mechanisms

and architecture of the genes involved in the adaptive phenotypic divergence (Rausher & Delph 2015), which can be studied by tracing the molecular selection patterns on the gene.

Most temperate amphibians exhibit a complex life-history where the larval stage is aquatic while the juvenile and adult stages are terrestrial. As fitness in amphibians is highly dependent on size at and timing of metamorphosis (e.g. Altwegg & Reyer 2003; Bradshaw *et al.* 2004; Earl & Whiteman 2015), larval development and growth strategies are key adaptations to environmental variation (e.g. Benard 2004; Rose 2005; Rowe & Ludwig 1991; Wilbur & Collins 1973). In a number of amphibian species, studies along latitudinal and altitudinal gradients have shown substantial phenotypic adaptation in larval growth and development to climatic variation (e.g. Berven & Gill 1983; Berven *et al.* 1979; Laugen *et al.* 2003; Luquet *et al.* 2015; Orizaola *et al.* 2010). In particular, low temperature and stronger time constraints at high latitudes and altitudes select for genetically faster larval development (Berven & Gill 1983; Laugen *et al.* 2003; Orizaola *et al.* 2010). However, apart from a few model species in other organisms than amphibians (e.g. Fabian *et al.* 2012), knowledge on the genetic factors affecting adaptive phenotypic variation along climatic gradients is almost entirely lacking (but see Johansson *et al.* 2013; Richter-Boix *et al.* 2013).

Genes involved in the thyroid hormone (TH) cascade are potential candidate genes for amphibian developmental rate (Johansson *et al.* 2013; Page *et al.* 2013; Richter-Boix *et al.* 2013; Voss *et al.* 2012; Voss *et al.* 2003), as they are involved in gene expression initiating morphological and developmental changes during metamorphosis (Buchholz *et al.* 2006; Das *et al.* 2009; Voss *et al.* 2003). Here we study the C/EBP-1 gene as a candidate gene with effects on larval development in the moor frog, *Rana arvalis*. This gene is a homologue to the mammalian C/EBP- α gene, which plays a general role in gene regulation for tissue differentiation, growth, cell cycle, and immune response (Tsukada *et al.* 2011). In amphibians, C/EBP-1 is an early response gene in the TH cascade (Chen & Atkinson 1997; Mochizuki *et al.* 2012). It encodes a tissue-specific transcription factor, which has a key function in reprogramming gene expression in the liver before and during amphibian metamorphosis for the necessary physiological changes involved in the transition from an aquatic to

a more terrestrial lifestyle (Chen & Atkinson 1997; Chen *et al.* 1994; Mochizuki *et al.* 2012). Richter-Boix *et al.* (2013; 2011) identified C/EBP-1 as a candidate gene for local adaptation in larval developmental time in *R. arvalis*. They demonstrated adaptive divergence in larval phenotypes among thermally contrasting habitats in a network of closely located wetlands, and linked this phenotypic variation to genetic variation at the C/EBP-1 gene, which appeared to be under directional selection imposed by local environmental conditions (Richter-Boix *et al.* 2013; Richter-Boix *et al.* 2011).

We investigated the genetic basis of larval development in the context of adaptation to large scale climatic variation in *R. arvalis*, with the specific aim to understand the role of variation at the C/EBP-1 gene in the timing of amphibian metamorphosis. As *R. arvalis* cannot overwinter as larvae, they have to complete their development before the ponds freeze in late autumn/early winter. This sets strict time constraints for larval development, and these constraints are much stronger at high latitudes as shown in other amphibians (Orizaola & Laurila 2016). Indeed, growth season length (number of days with mean temperature $\geq 5^{\circ}\text{C}$, defining the potential activity period) varies from ca. 278 days in the Lower Saxony region in northern Germany to ca. 161 days in the Norrbotten region in northern Sweden, selecting for faster development and growth in the north (Luquet *et al.* 2019).

Like many other species (Hewitt 2004), *R. arvalis* colonized Scandinavia via two different post-glacial colonization routes after the last glaciation (Babik *et al.* 2004; Cortázar-Chinarro *et al.* 2017; Knopp & Merila 2009). The southern route leads to Sweden from the south via Germany whereas the northern route reaches northern Sweden via Finland (Fig. 1). The contact zone between the two lineages is situated in northern central Sweden (between Uppsala and Umeå; Cortázar-Chinarro *et al.* 2017; Knopp & Merila 2009; Luquet *et al.* 2019).

We combined C/EBP-1 sequence data with neutral SNPs and common garden phenotypic data from five regions along a 1700 km latitudinal gradient from northern Germany to northern Sweden. This gradient includes the endpoints of both the southern and northern post-glacial colonization routes

into Scandinavia. We used a variety of methods to conduct a multifaceted analysis on C/EBP-1 and traced signals of current and historical selection, the footprint of colonization and associations with phenotypic variation on this candidate gene. Specifically, our aims were (1) to characterize patterns of genetic variation and differentiation at the C/EBP-1 gene along the latitudinal gradient, (2) to uncover signals of selection acting on C/EBP-1 among the different environments, (3) to examine whether variation at C/EBP-1 is linked to larval phenotypic variation, using genotype-phenotype association analyses.

Methods

R. arvalis is a common anuran amphibian in northern Europe with a distribution spanning from Croatia to northern Sweden (Sillero *et al.* 2014). We collected freshly-laid *R. arvalis* eggs in five regions along the latitudinal cline from Lower Saxony in northern Germany in the south to Norrbotten region in northern Sweden (Fig. 1, Table S1). Eggs were collected at the beginning of the breeding season at three sites per region, with an average distance of 20 km (8 to 50 km) among sites within regions. The collection sites were open-canopy wetlands situated in landscapes of mixed agricultural land and forest. At each site, we collected approximately 50 eggs from each of 13-30 freshly laid clutches (hereafter families) (see Table S1). The eggs were transported to Uppsala University, and raised to hatching in a climate room at constant 16°C and 16L:8D light rhythm.

DNA extraction, candidate gene amplification and preparation for sequencing

One individual per family was used for genetic analyses. These tadpoles were euthanized at Gosner stage 25 (Gosner 1960) with an overdose of MS222, preserved in 96% ethanol and stored at 4°C until DNA extraction. Total DNA was extracted from 399 individuals [26.6 ± 4.4 individuals per collection site, Table S1] using a high salt extraction protocol modified from Paxton (1996), with an additional ethanol precipitation step to purify DNA extracts from potential remnant inhibitors. Purity and

concentration of the DNA extracts were assessed with agarose gel electrophoresis and NanoDrop spectrophotometry.

We amplified and sequenced a 1028bp fragment of the C/EBP-1 gene, comprising the full 882bp protein coding part of the gene, and parts of the adjacent 5'- and 3'UTR regulatory regions. PCR amplifications were performed in three widely overlapping fragments of 582, 562 and 559bp in size, respectively, and sequenced on an Illumina MiSeq (for primer information see Table S2). PCR fragment size of app. 600bp was chosen according to the Illumina MiSeq kit v3 with a read length of 2 x 300 bases paired-end. Primers were designed based on *R. arvalis* C/EBP-1 sequences published in Richter-Boix et al. (2013) and *Lithobates (Rana) catesbeiana* RcC/EBP-1 gene (GenBank Accession U08604). All PCR primers were modified for Illumina sequencing at the 5' end with an individual 8bp barcode and a sequence of three N (see Table S3 for barcode information). PCR reactions were conducted in 25 µl volumes containing 1 µl DNA, 0.5 µM of each primer, 0.2 mM dNTP, 1x Phusion HF buffer, 0.1x Phusion GC buffer and 0.5 U Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific, Sweden) in deionized water. Reactions were initially denatured at 95°C for 3 min, followed by 35 cycles of denaturation at 95°C for 30 sec, annealing at the respective temperature for 30 sec (Table S2), and extension at 72°C for 45 sec. Reactions were ended by a final extension of 3 min at 72°C and cooled down to 8°C. To avoid contaminations, filter tips were used in all pre PCR procedures, pre and post PCR procedures were carried out in separate rooms, and negative controls were included in all amplifications.

PCR products were run on a 1.5% agarose gel using gelgreen (Biotium), quantified by eye, and 6-10 individual PCR products with similar concentrations were pooled to reduce the number of samples for subsequent purification. These sample pools were run on 1.5% agarose gels, the target band was excised from the gel and extracted using the MinElute Gel Extraction Kit (Qiagen® Sollentuna, Sweden). The concentration of each sample pool was measured with Quant-iT PicoGreen dsDNA assay kit (Invitrogen Life Technologies, Stockholm, Sweden) on a fluorescence microplate reader (Ultra 384; Tecan Group Ltd., Männedorf, Switzerland). The sample pools were then combined to

eight equimolar library pools, and libraries were prepared using the Illumina Truseq DNA PCR-Free Sample preparation kit (Illumina Inc, San Diego, CA). The eight libraries were combined into one Miseq300 run, and amplicon sequencing was carried out at the SNP&SEQ Technology Platform hosted at SciLifeLab in Uppsala.

Miseq sequence analyses

The raw Illumina Miseq reads were filtered by quality score using the standard Fastq filters, and demultiplexed using the FASTX-Toolkit 0.0.12 (http://hannonlab.cshl.edu/fastx_toolkit/) and a custom made script. Following the demultiplexing, barcodes and positions with a quality per base below 15 were trimmed off using Trimmomatic (Bolger *et al.* 2014). In addition, all reads shorter than 36 bp were disregarded. Sequences were aligned to the *L. catesbeiana* RcC/EBP-1 gene (GenBank Accession U08604) and genotypes were called using Geneious 6.1.8 (Kearse *et al.* 2012). Positions with a minimum coverage of 20 reads within an individual, and a minimum quality of 15 were considered, and SNPs with a within-individual minor allele frequency of >20% were called. For heterozygous sequences, we determined both C/EBP-1 haplotypes using the PHASE algorithm in DnaSP version 5.10.05 (Librado & Rozas 2009), which uses a Bayesian approach to infer haplotypes from genotypic data at linked loci from population genetic data (Stephens & Donnelly 2003; Stephens *et al.* 2001).

A number of samples revealed low coverage starting at position 774 and stretching across 10 to 150 bp in length. This drop in coverage is most likely the result of a considerable quality drop towards the end of the reads in combination with a general lower sequencing coverage of the third PCR fragment. Therefore, C/EBP-1 sequences for all individuals were truncated after position 773. Moreover, we removed the first 71bp of the sequences from the downstream analyses, because they were 5'UTR and invariable, leaving a 702 bp sequence.

Genetic analyses

DnaSP version 5.10.05 (Librado & Rozas 2009) was used to calculate sequence diversity statistics for the C/EBP-1 gene, and to calculate Tajima's D based on the total number of mutations (Tajima 1989). We used Arlequin 3.5.1.3 (Excoffier & Lischer 2010) to calculate observed and expected heterozygosity, and Fstat (Goudet 2001) was applied for calculating F_{IS} and allelic richness rarified to the smallest sample size. We measured population differentiation using the differentiation statistics F_{ST} (Wright 1965) and G' (Hedrick 2005). As the classical F_{ST} is at least partially dependent on the number of alleles within a population and can be influenced by heterozygosity, it can lead to difficulties in comparison among different loci (Meirmans & Hedrick 2011). Here, alternative statistics should be used in addition to F_{ST} . G'_{ST} is standardized by the maximum value of G_{ST} (F_{ST}) given the observed within-population diversity. However, Meirmans & Hedrick (2011) suggested to always present the original F_{ST} value because of its wide and long-standing use. F_{ST} -values for C/EBP-1 were calculated using a software developed by Neff and Fraser (2010) that allows for calculating confidence intervals for a single locus via resampling. As this resampling-technique is mainly used for small sample sizes (Neff & Fraser 2010), we used randomized subsets of the large sample sizes in the comparison between south and north, and averaged the results. G'_{ST} for the C/EBP-1 gene was calculated using the r-package *diveRsity* (Keenan *et al.* 2013). Confidence intervals were obtained by 999 bootstrap replicates.

MEGA 6.06 (Tamura *et al.* 2011) was used for sequence alignments and to calculate the dN/dS ratio on the whole alignment. We calculated the rates of non-synonymous (dN) and synonymous (dS) nucleotide substitutions using the Nei-Gojobori method with the Jukes Cantor correction, and tested for significant deviation from neutrality via Z-tests. We also tested for positive selection on specific codon sites along the alignment by estimating ω (=dN/dS) per codon position using OmegaMap (Wilson & McVean 2006) and the CODEML method (Yang 2007). OmegaMap uses a Bayesian approach including an estimation of recombination to identify positive selection across DNA sequences (Wilson & McVean 2006). We based the calculations on the 38 identified alleles

(disregarding allele frequencies), and ran the simulations twice for 500 000 iterations thinned for the posterior distribution every 1000 iterations. We used the inverse prior (0.010, 100) for ω and ρ , and the variable model allowing for ω and ρ to vary across codons, with an average block size of 5 and 15, respectively. Means and CI were generated with the 'summarize' module. As a comparison to OmegaMap, we tested for selection signals on specific codon sites using the method M8 from CODEML (Yang 2007) implemented in SELECTON (Stern *et al.* 2007). We ran the M8 model twice, once using the best phylogenetic tree for our data according to PhyML 3.0 (Guindon *et al.* 2010) under the HKY85 model of nucleotide substitution, and once using a star-shaped phylogenetic tree (i.e. "allele1, allele2, allele3, allele4, allele5, etc."). With this second, star-shaped tree we aimed to base the calculation on the assumption that all sequences are phylogenetically equally related. With this combined approach we aimed to disentangle if a positively selected site was indeed the result of selection or a result of an early split in the phylogeny of the sequences, i.e. many alleles carry a mutation in that specific position because the ancestral allele carried that mutation. If the positively selected site was the results of selection, it will be identified using both trees, whereas if it was due to an early split in phylogeny, the position will only be identified using the star-shaped tree. Identified positively selected codon positions were tested for associations with latitude using the Spearman rank correlation.

We constructed a minimum spanning network (Bandelt *et al.* 1999) of the 38 C/EBP-1 alleles using PopART v 1.7 (<http://popart.otago.ac.nz>). As PopART does not consider deletions as informative positions, we replaced deletions in six of the 38 C/EBP-1 alleles artificially by nucleotide substitutions for the purpose of network construction, assuming a single mutational step per deletion. This was done for illustration only, as deletions and nucleotide substitutions are different classes of mutations and do not evolve in the same way.

Neutral SNP variation

We used a subset of 1000 ddRAD generated putatively neutral SNP markers to estimate neutral genetic variation along the gradient. For a detailed description of the ddRAD procedure, see the Supplementary information. We calculated F_{ST} values based on the neutral SNPs in GENEPOP (Raymont & Rousset 1995) and confidence intervals were calculated using the Hierfstat R package (Goudet 2005), G'_{ST} was calculated using the r-package diveRsity (Keenan *et al.* 2013) with confidence intervals obtained by 999 bootstrap replicates.

Phenotypic analyses

We conducted a common garden experiment on tadpoles to investigate how development rates vary along the latitudinal gradient. For logistic reasons, phenotypic data from the Lower Saxony sites 2 and 3 were not available. When the tadpoles reached stage 25 (independent feeding; Gosner 1960), we haphazardly selected six tadpoles from each of eight to ten egg families per collection site. In total, 756 tadpoles from 126 families and 13 sites were included in the common garden experiment. The tadpoles were raised individually until metamorphosis in 0.75L opaque plastic containers and fed chopped spinach *ad libitum*. The water was changed every third day. The experiment was conducted in a walk-in climate controlled room with 19°C water temperature (corresponding to temperatures during late larval period; Richter-Boix *et al.* 2015) and 16L:8D photoperiod (corresponding to the light rhythm in Uppland region in May). When tadpoles reached metamorphosis (Gosner stage 42; emergence of the first forelimb), we recorded the number of days from the start of the experiment to metamorphosis (larval period) and mass at metamorphosis.

We compared the larval life-history traits (larval period, metamorphic mass) along the latitudinal gradient using linear mixed models with restricted maximum likelihood estimation computed in the R package lme4. In these models, region was included as a fixed factor, collection site and family were considered as random intercept effects. Family was nested within collection site and collection site

was nested within region. Significance of fixed effects was tested using a Kenward-Roger modification of F-tests (R package pbrtest, Halekoh & Hojsgaard 2014). Significance of random effects was tested using likelihood ratio test comparing models with or without the tested random effect in the full fixed effect structure.

Genotype - phenotype associations

We used the C/EBP-1 alleles for genotype-phenotype associations rather than SNPs within the gene, because all variable positions in a candidate gene are highly linked due to their physical proximity within a few hundred base pairs and cannot be treated as independent markers. Using alleles rather than SNPs can provide a more powerful strategy to test for phenotypic associations (Vasemägi & Primmer 2005).

We used two different methods to explore the relationship between the C/EBP-1 genotype and larval phenotypic variation in the 113 families for which both genetic and phenotypic data were available. In the first approach, to avoid the problem of multicollinearity among different C/EBP-1 alleles in a regression analysis, we performed a multiple correspondence analysis (MCA) to combine the genotypic information into synthetic variables. We excluded low frequency alleles (<3%), because they occurred in three or less individuals. Each individual was thus genetically characterized by the presence or absence of the five most frequent C/EBP-1 alleles (A01, A02, A03, A04, A06), and a MCA was performed on these data (Tenenhaus & Young 1985). Subsequently, the first two MCA axes were used as predictors in analyses of variance models explaining larval phenotypes (larval period and metamorphic mass). MCA were calculated in the ade4 R package (Dray & Dufour 2007).

Second, we investigated the effects of the C/EBP-1 gene on phenotypic variation using the NOIA R package (Alvarez-Castro & Carlborg 2007), a statistical framework for quantifying genetic effects on phenotypic traits. NOIA uses linear and multilinear regression and discriminates between additive genetic effects and dominance effects. Here, the correlation between genotype and phenotype was

analyzed for each C/EBP-1 allele separately, based again on alleles occurring at >3% frequency (A01, A02, A03, A04, A06), and p-values were Bonferroni-corrected for multiple testing. As phenotypic data, we used the family means and matched these data with the genotyped individual of the same family. Due to the different routes of post-glacial colonization into Scandinavia, *R. arvalis* from different regions of the gradient differ in demographic history and genetic background (Cortázar-Chinarro *et al.* 2017). To reduce the potential effect of differences in genetic backgrounds, we performed separate analyses within northern and southern datasets, the northern comprising 54 families from Västerbotten and Norrbotten, and the southern 59 families from Lower Saxony, Skåne and Uppland.

Results

Genetic diversity at the C/EBP-1 gene and variation along the gradient

We analyzed 702 bp of the coding region of C/EBP-1 in 364 individuals from 15 collection sites. Another 35 individuals were excluded from analyses, because they did not reveal sufficient data due to failure in either PCR amplification or sequencing. We found 38 C/EBP-1 nucleotide alleles, which translated into 11 unique amino acid sequences (Table 1, Fig. 1). In total, there were 18 variable nucleotide positions, of which six were non-synonymous and 12 synonymous substitutions (Figure S1). None of the alleles contained premature stop-codons, but six alleles contained deletions that maintained the reading frame and were therefore considered functional. Average pairwise nucleotide differences among alleles (theta k), number of segregating sites and nucleotide diversity were relatively low, indicating that large parts of the sequences were identical among alleles (Table 1).

Among geographical regions, genetic variation differed markedly, with the highest diversity in Lower Saxony in the southernmost part of the gradient. Number of alleles, allelic richness, number of private alleles and observed heterozygosity were highest in the south and declined towards the

northern regions (Table 1, Fig. 1). The lowest value for heterozygosity was observed in the northernmost region Norrbotten, with one dominant allele and >93% homozygous individuals. Allelic richness and number of alleles were lowest in the Västerbotten region (Table 1). Genetic differentiation among the five regions was very high, with a global G'_{ST} of 0.793 (global $F_{ST} = 0.307$). While G'_{ST} was comparably low among the three southern regions (0.127 - 0.356, Table 2) and between the two northern regions ($G'_{ST} = 0.376$), divergence was strong among the southern and the northern regions (pairwise $G'_{ST} = 0.821 - 0.997$, Table 2; see Table S4 for the corresponding F_{ST} comparisons).

Among the 38 identified C/EBP-1 alleles, 25 were private within a single region, and 13 alleles were shared among at least two regions along the gradient (Fig. 1). Among these shared alleles, 10 were restricted to the three southern regions of the gradient (Lower Saxony, Skåne, Uppland). Only three alleles were shared between the south and the north, alleles A01 (Lower Saxony, Västerbotten, Norrbotten), A02 (all regions) and A04 (Lower Saxony, Skåne, Uppland, Västerbotten; Fig. 1).

The C/EBP-1 alleles present in the two northern regions showed less sequence divergence as compared to the alleles present in the southern regions. In a minimal spanning network of the alleles (Fig. 2), the alleles present in the north are not scattered across the whole network, but cluster close to each other, indicating high sequence similarity. This similarity is also indicated by the lower values for theta k in the two northern regions (Table 1).

Signature of selection on the C/EBP-1 gene

We tested whether variation at C/EBP-1 along the gradient was shaped by contemporary selection via comparing G'_{ST} on C/EBP-1 to a set of 1000 putatively neutral genome wide SNP markers (Table 2). In all but one the pairwise comparison among the regions, G'_{ST} for the C/EBP-1 gene was significantly higher than for the neutral SNPs (no overlap between confidence intervals for the two different sets of markers, Table 2), suggesting divergent selection on C/EBP-1 along the whole

gradient. An exception was the pairwise comparison between Skåne and Uppland regions, where G'_{ST} was slightly lower for the C/EBP gene, however, this difference was not significant. The pattern of divergent selection on the candidate gene was most pronounced between south and north (Table 2).

We also tested for signals of historical selection on the C/EBP-1 gene by calculating the dN/dS ratio (ω) and Tajima's D. The dN/dS ratio revealed an excess of synonymous substitutions in all regions, leading to significant signals of purifying selection in the total sample and the three southern regions (Table 3). In the two northern regions, statistical power for the detection of selection signals was lower due to the low number of alleles present. The negative values for Tajima's D in the total sample and three of the regions suggest purifying selection having acted on C/EBP-1 (or a recent population expansion), however, none of these values was significant (Table 3).

Looking more into detail by estimating ω per codon position along the alignment, using OmegaMap and CODEML, we found that ω varies substantially among positions (Fig. 3). The vast majority of codon positions was under purifying selection (ω close to zero), indicating that most parts of the gene are structurally conserved and invariable. There was, however, significant indication for positive selection on codon positions 61 and 77, indicating that the C/EBP-1 alleles diverge especially in these positively selected amino acid positions. While position 77 was detected by all three models for positive selection, position 61 was only identified when phylogenetic information was not pre-defined (Fig. 3a, b and c). This indicates that variation at position 61 may be attributed to an early split in the phylogeny of C/EBP-1 alleles, whereas position 77 shows positive selection independent of phylogeny.

Position 61 was either the amino acid glutamate or aspartate. While aspartate was the predominant amino acid in individuals from higher latitudes, glutamate was dominating in individuals from lower latitudes ($r_{s(362)} = 0.762$, $p < 0.0001$). Variation at position 77, in contrast, was formed by two adjacent SNPs (see Figure S1) and resulted in three different amino acids at this position: alanine, serine and threonine. The presence of alanine was negatively ($r_{s(362)} = -0.233$, $p < 0.0001$) and serine positively

($r_{S(362)} = 0.188$, $p = 0.0003$) correlated with latitude, while the presence of threonine was not correlated with latitude ($r_{S(362)} = -0.005$, $p = 0.92$).

Phenotypic variation

In the common garden, length of larval period and mass at metamorphosis significantly differed along the latitudinal gradient (larval period: $F_{4, 7.995} = 22.973$, $p = 0.0002$, mass: $F_{4, 7.947} = 57.802$, $p < 0.0001$; Fig. 4). The divergence strongly depended the geographic region, with tadpoles from the two northern regions showing a 21 % shorter larval period (ca. 9 days) and being 27 % heavier than tadpoles from the three southern regions (for complete pairwise contrasts see Table S5; Fig. 4). Random collection site effect was only significant for larval period and random family effect was significant for all larval traits (Table S6).

Genotype - phenotype associations

To associate genotypic variation at the C/EBP-1 gene with the phenotypic data, we performed a multiple correspondence analysis based on the presence or absence of the five most common C/EBP-1 alleles and then included the first two MCA axes as predictors of phenotypes in ANOVA models. The first MCA axis explained 23 % of the allelic variation in C/EBP-1 and sorted the individuals into two clusters corresponding to the northern and southern colonization routes: Norrbotten and Västerbotten on the left versus Lower Saxony, Skåne and Uppland on the right (Fig. 5A). Individuals carrying A02, A03 or A06 reached positive values along the first MCA axis (Lower Saxony, Skåne and Uppland), and individuals carrying A01 or A04 reached negative values (Norrbotten and Västerbotten, Fig. 5B). MCA2, on the other hand, explained 17 % of the variation and split the northernmost region Norrbotten from Västerbotten and the southern regions, thereby separating the A01 homozygote individuals (Norrbotten) from A01 heterozygotes and the presence of A04 (Västerbotten). Thus, MCA2 shows how the northernmost region Norrbotten genetically stands out

against the other regions, and additionally differentiates between the two regions of the northern colonization route.

The ANOVA models revealed that allele distribution along the two MCA axes significantly explains phenotypic variation (Table 4): individuals with high MCA1 scores (southern regions) showed a longer larval period and a lower mass at metamorphosis than individuals with a low MCA1 score (northern regions, Table 4). MCA2 also revealed significant effects on phenotypic variation, as individuals with high MCA2 scores showed a shorter larval period (Table 4). However, MCA1 and 2 did not explicitly illustrate variation at the C/EBP-1 gene, but also other genetic variation along the gradient and the dual colonization routes, and this pattern may not be causal.

We also analysed genotype-phenotype associations with NOIA separately within southern and northern ends of the gradient (Table 5). We found no evidence for dominance effects, and all associations reported below are additive. When considering only the northern regions, A01 was associated with shorter larval period and a lower mass at metamorphosis, while A04 was associated with a longer larval period and a higher mass at metamorphosis (Table 5a). There was a non-significant trend ($p = 0.064$, data not shown) for shorter larval period in individuals with A01 when performing the analysis within the Västerbotten region only. Within the southern regions, A06 was associated with a shorter larval period.

Discussion

In this study, we investigated the C/EBP-1 gene as a candidate gene for larval development, and analyzed associations between sequence variation and phenotypic variation. Along the 1700 km latitudinal gradient, we found poleward decline in genetic diversity and that the distribution of C/EBP-1 alleles along the gradient reflected the dual colonization route into Scandinavia, with a notable shift in allelic composition between the regions closest to the contact zone between the colonization lineages (Uppland and Västerbotten). We found signals of divergent contemporary

selection acting on the C/EBP-1 gene, and analyses of historical selection indicated that while C/EBP-1 has predominantly been under purifying selection, some positions have experienced positive selection. We also found significant phenotype-genotype associations, as specific C/EBP-1 alleles were correlated with larval development demonstrating a link between allelic variation in the gene and fitness-related phenotypes. We note that while we cannot totally exclude that the phenotype-genotype associations can to some extent be influenced by the genetic background due to the dual post-glacial colonization lineages, our results are reinforced by analyses taking into account the dual lineages. Taken together, our results suggest that at a large-scale climatic background C/EBP-1 is a candidate gene for climate-mediated adaptation in larval development in *R. arvalis*.

Genetic variation, poleward decline and recolonization

Overall, we found a relatively high number of 38 alleles at the C/EBP-1 gene in 364 individuals. Earlier, Richter-Boix et al. (2013) identified 36 C/EBP-1 alleles in 379 individuals from divergent habitats within a small geographical area in central Sweden, and one might have expected an even higher number of alleles in our study, considering the large geographical range of our sampling. However, Richter-Boix et al. (2013) sequenced the full coding region of 882bp, whereas we were not able to include in our analyses the last 180bp containing some polymorphic sites. Furthermore, the earlier study included a greater variety of breeding habitats than the present study, which may also have increased the number of alleles found in Richter-Boix *et al.* (2013). Consequently, it seems possible that our study may somewhat underestimate the actual variation at C/EBP-1.

We found a strong decline in genetic variation in C/EBP-1 from the more central southern populations towards the northern limit of the species distribution. According to the central-marginal hypothesis, populations in the periphery of the species range are expected to exhibit lower genetic diversity and greater genetic differentiation relative to the core populations as a consequence of smaller effective population size and greater geographical isolation (Eckert *et al.* 2008). A decline in

genetic variation at higher latitudes can also be created by historical colonization processes, for instance, by founder effects during post-glacial range expansions (Eckert *et al.* 2008; Hewitt 2000). Disentangling the effects of sequential post-glacial founder events from the central-marginal effects of population isolation and small effective population size is difficult as both factors often positively correlate with latitude (Eckert *et al.* 2008).

Along the southern route, number of alleles, number of private alleles, allelic richness and observed heterozygosity in C/EBP-1 decreased towards Uppland, the northernmost region of this route in our sampling scheme. A similar pattern has been found using different genetic markers in *R. arvalis* (microsatellite and MHC variation; Cortázar-Chinarro *et al.* 2017; Knopp & Merila 2009) and other species (Hewitt 2000) in Scandinavia. Along the northern route, all measures of genetic variation in C/EBP-1 - except heterozygosity - were higher in Norrbotten than in Västerbotten, in accordance with the hypothesis that Norrbotten region was colonized earlier than Västerbotten. Heterozygosity measures, however, did not follow the same trend as the by far lowest values were detected in Norrbotten. We suggest that selection for fast development due to the strong time constraint in Norrbotten can explain the low heterozygosity in C/EBP-1. This notion is indirectly supported by the patterns of phenotypic variation and genotype-phenotype associations (see below). Importantly, it seems possible that the limited genetic variation in the northern populations may limit the capacity of these populations to adapt to changing climate. This may be the case especially in Norrbotten populations with very fast development rate, which may have to rely on gene flow from southern populations to obtain alleles linked with slower development likely selected for under a warmer climate (see below).

We observed very little allelic overlap between the south and the north in our study, as only three out of 38 C/EBP-1 alleles occurred in south and north, and only one C/EBP-1 allele was shared by all five regions. This observation fits well with the demographic history of two post-glacial recolonization routes into Scandinavia. The shared C/EBP-1 alleles might be shared ancestral alleles, remnants from an ancestral source population that had been passed on to both the southern and the northern

colonizers, or these alleles might have introgressed from one colonization lineage to the other at secondary contact in Scandinavia.

Divergent selection patterns on C/EBP-1 gene and evidence for function

Three lines of evidence in this study support C/EBP-1 as a target of natural selection and of functional importance. First, we detected that differentiation among regions at C/EBP-1 was significantly higher compared to neutral markers, indicating recent divergent selection on the gene. Second, we also demonstrated significant historical selection on the gene. Third, we showed that variation at C/EBP-1 is associated with fitness-related larval phenotypes across a large spatial scale. In the earlier study, Richter-Boix *et al.* (2013) demonstrated that variation in C/EBP-1 was associated with variation in small-scale variation in larval development among contrasting thermal environments. Identifying the same gene in independent cases of local adaptation is strong evidence for functional importance of the gene (Conte *et al.* 2012; Tiffin & Ross-Ibarra 2014).

Generally, in systems where the selection gradient is one-dimensional, such as altitudinal or latitudinal clines, drift and migration can generate patterns across the environmental gradients that closely mimic adaptive clines (Vasemägi & Primmer 2005). For instance, northward expansion from a glacial refugium may generate clines in neutral allele frequencies from south to north, which coincide with climatic gradients, such as temperature (Hoban *et al.* 2016). In these systems, signals of selection and patterns derived from demographic processes are overlaid and hard to disentangle. However, in the present study we compared genetic patterns derived from the C/EBP-1 gene with neutral genetic markers, shaped by drift and demographic history only, enabling us to clearly see that the C/EBP-1 gene carries the footprint of natural selection.

G'_{ST} values on the C/EBP-1 gene were higher than G'_{ST} in the neutral SNPs indicating divergent contemporary selection on the C/EBP-1 gene along the whole gradient, the sole exception being divergence between Skåne and Uppland regions. While divergent selection was evident within both

lineages, it was especially pronounced between the south and the north. We assume that this pattern is driven by the more extreme environmental conditions in the north, mediated by selection for faster development and growth close to the northern range limit of the species (see below and Laugen et al. 2003; Palo et al. 2003, Lindgren & Laurila 2009 for similar results in *R. temporaria*), although colonization routes can also play a role. The signature of historical selection on the C/EBP-1 gene revealed that the largest part of the gene had been under purifying selection, indicating that the gene is conserved and under strong functional constraint. In general, strong selection on stability and structural constancy of the gene function is expected for a gene involved in a highly complex process such as larval development. Variation among the different C/EBP-1 alleles was mainly concentrated at two amino acid positions, 61 and 77, which showed significant indication of positive selection among alleles. Variation at these positions was also correlated with latitude, so that alleles predominantly from the north carried aspartate at position 61 and serine at position 77 (for instance, A01), whereas alleles from the south carried predominantly glutamate at position 61, and alanine at position 77 was exclusively found in southern alleles. This pattern of positive selection of specific amino acids combined with the latitudinal associations suggests that these two positions are functionally involved in the adaptation to latitudinal variation. However, as variation at position 61 seemed to have arisen from an early split in the phylogeny of C/EBP-1 alleles, the latitudinal pattern at this position might also reflect the different demographic histories and the dual recolonization routes.

Phenotypic patterns along the gradient and genotype-phenotype associations

We found that tadpoles from the two northernmost regions developed faster, and were larger at metamorphosis than tadpoles from the three southern regions. These results agree with previous studies on amphibians (e.g., Berven & Gill 1983; Laugen *et al.* 2005; Laugen *et al.* 2003; Lindgren & Laurila 2009; Orizaola *et al.* 2010) and many other ectotherm species (reviewed by Conover *et al.* 2009; Dmitriew 2011). In general, faster development is favored in more time-constrained

environments, whereas large size at metamorphosis may increase overwintering success during the long winter (e.g., Altwegg & Reyer 2003; Munch *et al.* 2003). The adaptive importance of the phenotypic divergence and strong natural selection in the present system was supported by F_{ST} - Q_{ST} analyses showing that the Q_{ST} for the three phenotypic traits along the latitudinal gradient was roughly four times higher than F_{ST} for neutral SNPs (Luquet *et al.* 2019).

The distribution of C/EBP-1 alleles was to a large degree sorted by the two different recolonization routes, as reflected by MCA1. We found several alleles significantly associated with phenotypic variation: individuals carrying A01 were mostly from the north and developed fast into large metamorphs. Individuals carrying A02, A03 or A06 were mostly from the south and developed slowly into small metamorphs, whereas individuals with A04 occurred in both the southern and northern part of the gradient and developed slowly. The second MCA axis loaded heavily on A01 and clearly separated Norrbotten from all other regions, suggesting that the more extreme climatic conditions in this region may have increased the occurrence of A01 in these populations, as also indicated by the negative relationship between MCA2 and larval period.

While we cannot exclude that the phenotype-genotype associations can be influenced by the different genetic background due to the dual post-glacial colonization lineages, the results from MCA were corroborated in the NOIA analyses which were conducted separately within colonizing lineages. Three of the five most common alleles (A01, A04 and A06) were associated with larval development. The predominantly northern allele A01 was strongly associated with fast development, and there was also evidence for association with lower mass at metamorphosis, suggesting a fast development strategy resulting in smaller size within the northern lineage. Also A06, found exclusively in the southern regions, was linked with faster development. This indicates that the most common allele associated with fast development was different between the southern and the northern lineages, as A01 occurred within the southern lineage only at low frequency in Germany.

Individuals carrying A04 developed slowly. Remarkably, while A04 allele was found in Västerbotten and in the southern regions, it did not occur in Norrbotten, suggesting either gene flow from the southern regions over the secondary contact zone, or that A04 allele has been selected out from the Norrbotten populations where selection for fast development can be expected to be strongest. As high development rates are inherently costly (Dmitriew 2011), selection favoring slow-developing phenotypes at the southern edge of the northern lineage seems feasible, possibly suggesting gene flow over the secondary contact zone. This would also explain the high frequency of A04 in Västerbotten populations, as the selective advantage in the absence of other slow-development alleles can be strong. On the other hand, the lack of A01 south of the contact zone in Uppland region may suggest that high development rate may directly be selected against in lower latitudes in line with studies on several ectotherms (Conover *et al.* 2009; Dmitriew 2011), including amphibians (Laurila *et al.* 2008; Laurila *et al.* 2006).

Interestingly, two of the allelic effects that we detected were confirmed in the earlier study of Richter-Boix *et al.* (2013). Allele HTRb13, which corresponds to A06 in our study, was associated with a short larval period, as we also found in the present study using both NOIA and MCA methods. Moreover, Allele HTRb9, which corresponds to A02 in our study, was associated with a long larval period, as was A02 in our MCA analysis. These identical genotype-phenotype associations replicated in different studies provide strong evidence for the causative functional role of the C/EBP-1 gene in local adaptation of larval development in *R. arvalis*.

Conclusions

In this study we traced signals of selection, local adaptation and the footprint of post-glacial colonization in a candidate gene for amphibian larval development. We showed evidence for divergent selection and functional importance of the C/EBP-1 gene in local adaptation of larval development in *R. arvalis*, although the complicating effects of the dual postglacial colonization

lineages have to be kept in mind in the latter case. Importantly, genetic variation in C/EBP-1 was much smaller in the northern populations than in the southern populations suggesting that the potential for adaptation is more limited in these populations. This was especially the case in the Norrbotten populations showing fastest development rates. While our results suggest that the ability of these populations to adapt to warming climate seems limited in the absence of gene flow from the south, they also suggest that introgression over the contact zone may have occurred. In general, this study highlights how the combination of candidate gene information, neutral genetic markers and phenotypic data provides a powerful approach for understanding evolutionary processes in response to spatial and temporal climatic variation.

Acknowledgements

We thank Deike Lüdtke and Fredrik Tegnér for help in maintaining the tadpoles, Cátia Chaves for helping with the molecular genetic work, and Peter Schweitzer and Qi Sun for their help with the script for demultiplexing the data. Björn Rogell provided constructive comments on the manuscript. The eggs were collected with permits from Authority for Nature Protection of Lower Saxony and the Lower Saxony region and respective county boards in Sweden. We are thankful to Anders Hallengren (Länstyrelsen Skåne), Anders Forsgren and Stefan Andersson (Piteå Kommun), Thomas Brandt (Ökologische Schutzstation Steinhuder Meer e.V), and Bernd Rittberg and Andreas Jacob (Authority for Nature Protection of Lower Saxony and the Hanover region) for helping with the sampling locations and permits. The experiment was conducted with permit C40/14 from the Ethical committee for animal experiments in Uppsala County. The study was financially supported by the Swedish Research Council (project 621-2013-4503 to AL), Carl Tryggers Stiftelse för Vetenskaplig Forskning (to YML, EL) and Zoologiska stiftelsen (YML, EL). Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life laboratory. The SNP&SEQ Platform is supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

Data accessibility

The C/EBP-1 nucleotide sequences are available at GenBank with accession numbers: MK749692 - MK749729. The data including C/EBP-1 and SNP genotypes and common garden phenotypes are available from the Dryad Digital Repository: doi:10.5061/dryad.2508t4h.

References

- Altwegg R, Reyer HU (2003) Patterns of natural selection on size at metamorphosis in water frogs. *Evolution* **57**, 872-882.
- Alvarez-Castro JM, Carlborg O (2007) A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. *Genetics* **176**, 1151-1167.
- Angilletta MJ (2009) *Thermal adaptation: a theoretical and empirical synthesis* Oxford University Press, Oxford.
- Babik W, Branicki W, Sandera M, *et al.* (2004) Mitochondrial phylogeography of the moor frog, *Rana arvalis*. *Molecular Ecology* **13**, 1469-1480.
- Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.
- Benard MF (2004) Predator-induced phenotypic plasticity in organisms with complex life histories. *Annual Review of Ecology, Evolution, and Systematics* **35**, 651-673.
- Berven KA, Gill DE (1983) Interpreting geographic variation in life-history traits. *American Zoologist* **23**, 85-97.
- Berven KA, Gill DE, Smith-Gill SJ (1979) Countergradient selection in the green frog, *Rana clamitans*. *Evolution* **33**, 609-623.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.
- Bonin A, Taberlet P, Miaud C, Pompanon F (2005) 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**, 773-783.
- Bradshaw WE, Zani PA, Holzapfel CM (2004) Adaptation to temperate climates. *Evolution* **58**, 1748-1762.
- Buchholz DR, Paul BD, Fu L, Shi Y-B (2006) Molecular and developmental analyses of thyroid hormone receptor function in *Xenopus laevis*, the African clawed frog. *General and Comparative Endocrinology* **145**, 1-19.
- Chen Y, Atkinson BG (1997) Role for the *Rana catesbeiana* homologue of C/EBP α in the reprogramming of gene expression in the liver of metamorphosing tadpoles. *Developmental Genetics* **20**, 152-162.
- Chen Y, Hu H, Atkinson BG (1994) Characterization and expression of C/EBP-like genes in the liver of *Rana catesbeiana* tadpoles during spontaneous and thyroid hormone-induced metamorphosis. *Developmental Genetics* **15**, 366-377.
- Conover DO, Duffy TA, Hice LA (2009) The covariance between genetic and environmental influences across ecological gradients. *Annals of the New York Academy of Sciences* **1168**, 100-129.
- Conte GL, Arnegard ME, Peichel CL, Schluter D (2012) The probability of genetic parallelism and convergence in natural populations **279**, 5039-5047.

- Cortázar-Chinarro M, Lattenkamp EZ, Meyer-Lucht Y, *et al.* (2017) Drift, selection, or migration? Processes affecting genetic differentiation and variation along a latitudinal gradient in an amphibian. *BMC Evolutionary Biology* **17**, 189.
- Das B, Heimeier RA, Buchholz DR, Shi Y-B (2009) Identification of Direct Thyroid Hormone Response Genes Reveals the Earliest Gene Regulation Programs during Frog Metamorphosis. *Journal of Biological Chemistry* **284**, 34167-34178.
- Dmitriew CM (2011) The evolution of growth trajectories: what limits growth rate? *Biological Reviews* **86**, 97-116.
- Dray S, Dufour A-B (2007) The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software* **22**, 1-20.
- Earl JE, Whiteman HH (2015) Are commonly used fitness predictors accurate? A meta-analysis of amphibian size and age at metamorphosis. *Copeia* **103**, 297-309.
- Eckert CG, Samis KE, Loughheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology* **17**, 1170-1188.
- Endler JA (1977) *Geographic variation, speciation, and clines* Princeton University Press, Princeton
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.
- Fabian DK, Kapun M, Nolte V, *et al.* (2012) Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Molecular Ecology* **21**, 4748-4769.
- Gosner KL (1960) A simplified table for staging anuran embryos and larvae with notes on identification. *Herpetologica* **16**, 183-190.
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from <http://www.unil.ch/izea/software/fstat.html>.
- Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* **5**, 184-186.
- Guindon S, Dufayard JF, Lefort V, *et al.* (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**, 307-321.
- Guo B, Lu D, Liao WB, Merilä J (2016) Genomewide scan for adaptive differentiation along altitudinal gradient in the Andrew's toad *Bufo andrewsi*. *Molecular Ecology* **25**, 3884-3900.
- Halekoh U, Hojsgaard S (2014) Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest. *Journal of Statistical Software* **59**, 1-32.
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* **59**, 1633-1638.
- Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**.
- Hewitt GM (2004) The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology* **1**, 4.
- Hoban S, Kelley JL, Lotterhos KE, *et al.* (2016) Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist* **188**, 379-397.
- Hoekstra HE, Hirschmann RJ, Bunday RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**, 101-104.
- Johansson F, Veldhoen N, Lind MI, Helbing CC (2013) Phenotypic plasticity in the hepatic transcriptome of the European common frog (*Rana temporaria*): the interplay between environmental induction and geographical lineage on developmental response. *Molecular Ecology* **22**, 5608-5623.
- Jones FC, Grabherr MG, Chan YF, *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature (London)* **484**, 55-61.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.
- Kearse M, Moir R, Wilson A, *et al.* (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

- Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* **4**, 782-788.
- Keller I, Alexander JM, Holderegger R, Edwards PJ (2013) Widespread phenotypic and genetic divergence along altitudinal gradients in animals. *Journal of Evolutionary Biology* **26**, 2527-2543.
- Knopp T, Merila J (2009) The postglacial recolonization of Northern Europe by *Rana arvalis* as revealed by microsatellite and mitochondrial DNA analyses. *Heredity* **102**, 174-181.
- Lamichhaney S, Barrio AM, Rafati N, *et al.* (2012) Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 19345-19350.
- Laugen AT, Laurila A, Jonsson KI, Soderman F, Merila J (2005) Do common frogs (*Rana temporaria*) follow Bergmann's rule? *Evolutionary Ecology Research* **7**, 717-731.
- Laugen AT, Laurila A, Rasanen K, Merila J (2003) Latitudinal countergradient variation in the common frog (*Rana temporaria*) development rates - evidence for local adaptation. *Journal of Evolutionary Biology* **16**, 996-1005.
- Laurila A, Lindgren B, Laugen AT (2008) Antipredator defenses along a latitudinal gradient in *Rana temporaria*. *Ecology* **89**, 1399-1413.
- Laurila A, Pakkasmaa S, Merilä J (2006) Population divergence in growth rate and antipredator defences in *Rana arvalis*. *Oecologia* **147**, 585-595.
- Lee YW, Gould BA, Stinchcombe JR (2014) Identifying the genes underlying quantitative traits: a rationale for the QTN programme. *AoB Plants* **6**.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451 - 1452.
- 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.
- Luquet E, Léna J-P, Miaud C, Plénet S (2015) Phenotypic divergence of the common toad (*Bufo bufo*) along an altitudinal gradient: evidence for local adaptation. *Heredity* **114**, 69-79.
- Luquet E, Rödin Mörch P, Cortazar-Chinarro M, *et al.* (2019) Postglacial colonization routes coincide with a life history breakpoint along a latitudinal gradient. *Journal of Evolutionary Biology*, in press.
- Meirmans PG, Hedrick PW (2011) Assessing population structure: FST and related measures. *Molecular Ecology Resources* **11**, 5-18.
- Meyer-Lucht Y, Luquet E, Johannesdottir F, *et al.* (2019) Data from: Genetic basis of amphibian larval development along a latitudinal gradient: gene diversity, selection and links with phenotypic variation in transcription factor C/EBP-1. Dryad Digital Repository. doi:10.5061/dryad.2508t4h
- Mochizuki K, Goda T, Yamauchi K (2012) Gene expression profile in the liver of *Rana catesbeiana* tadpoles exposed to low temperature in the presence of thyroid hormone. *Biochemical and Biophysical Research Communications* **420**, 845-850.
- Munch SB, Mangel M, Conover DO (2003) Quantifying natural selection on body size from field data: Winter mortality in *Menidia menidia*. *Ecology* **84**, 2168-2177.
- Neff BD, Fraser BA (2010) A program to compare genetic differentiation statistics across loci using resampling of individuals and loci. *Molecular Ecology Resources* **10**, 546-550.
- Orizaola G, Laurila A (2016) Developmental plasticity increases at the northern range margin in a warm-dependent amphibian. *Evolutionary applications* **9**, 471-478.
- Orizaola G, Quintela M, Laurila A (2010) Climatic adaptation in an isolated and genetically impoverished amphibian population. *Ecography* **33**, 730-737.
- Page RB, Boley MA, Kump DK, Voss SR (2013) Genomics of a metamorphic timing QTL: met1 maps to a unique genomic position and regulates morph and species-specific patterns of brain transcription. *Genome biology and evolution* **5**, 1716-1730.

- Paxton RJ, Thorén PA, Tengö J, Estoup A, Pamilo P (1996) Mating structure and nestmate relatedness in a communal bee, *Andrena jacobi* (Hymenoptera, Andrenidae), using microsatellites. *Molecular Ecology* **5**, 511-519.
- Poelstra JW, Vijay N, Bossu CM, *et al.* (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**, 1410-1414.
- Rausher MD, Delph LF (2015) Commentary: When does understanding phenotypic evolution require identification of the underlying genes? *Evolution* **69**, 1655-1664.
- Raymont M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism *Journal of Heredity* **86**, 248-249.
- Richter-Boix A, Quintela M, Kierczak M, Franch M, Laurila A (2013) Fine-grained adaptive divergence in an amphibian: genetic basis of phenotypic divergence and the role of nonrandom gene flow in restricting effective migration among wetlands. *Molecular Ecology* **22**, 1322-1340.
- Richter-Boix A, Quintela M, Segelbacher G, Laurila A (2011) Genetic analysis of differentiation among breeding ponds reveals a candidate gene for local adaptation in *Rana arvalis*. *Molecular Ecology* **20**, 1582-1600.
- Richter-Boix A, Katzenberger M, Duarte H, *et al.* (2015) Local divergence of thermal reaction norms among amphibian populations is affected by pond temperature variation. *Evolution* **69**, 2210-2226.
- Rose C (2005) Integrating ecology and developmental biology to explain the timing of frog metamorphosis. *Trends in Ecology & Evolution* **20**, 129-135.
- Rowe L, Ludwig D (1991) Size and timing of metamorphosis in complex life cycles: time constraints and variation. *Ecology* **72**, 413-427.
- Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**, 807-820.
- Sillero N, Campos J, Bonardi A, *et al.* (2014) Updated distribution and biogeography of amphibians and reptiles of Europe. *Amphibia-Reptilia* **35**, 1-31.
- Stephens M, Donnelly P (2003) A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data. *The American Journal of Human Genetics* **73**, 1162-1169.
- Stephens M, Smith NJ, Donnelly P (2001) A New Statistical Method for Haplotype Reconstruction from Population Data. *The American Journal of Human Genetics* **68**, 978-989.
- Stern A, Doron-Faigenboim A, Erez E, *et al.* (2007) Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach *Nucleic Acids Research* **35**, 506-511.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.
- Tamura K, Peterson D, Peterson N, *et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**, 2731-2739.
- Tenenhaus M, Young FW (1985) An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika* **50**, 91-119.
- Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution* **29**, 673-680.
- Tsukada J, Yoshida Y, Kominato Y, Auron PE (2011) The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation. *Cytokine* **54**, 6-19.
- Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology* **14**, 3623-3642.
- Voss SR, Kump DK, Walker JA, Shaffer HB, Voss GJ (2012) Thyroid hormone responsive QTL and the evolution of paedomorphic salamanders. *Heredity* **109**, 293-298.

- Voss SR, Prudic KL, Oliver JC, Shaffer HB (2003) Candidate gene analysis of metamorphic timing in ambystomatid salamanders. *Molecular Ecology* **12**, 1217-1223.
- Wilbur H, Collins J (1973) Ecological aspects of amphibian metamorphosis. *Science* **182**, 1305-1314.
- Wilczek AM, Roe JL, Knapp MC, *et al.* (2009) Effects of Genetic Perturbation on Seasonal Life History Plasticity. *Science* **323**, 930-934.
- Wilson DJ, McVean G (2006) Estimating Diversifying Selection and Functional Constraint in the Presence of Recombination. *Genetics* **172**, 1411-1425.
- Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19**, 395-420.
- Yang W, Qi Y, Bi K, Fu J (2012) Toward understanding the genetic basis of adaptation to high-elevation life in poikilothermic species: a comparative transcriptomic analysis of two ranid frogs, *Rana chensinensis* and *R. kukunoris*. *BMC Genomics* **13**, 588.
- Yang ZH (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**, 1586-1591.

YML, EL and AL conceived and planned the study. YML, EL, FJ and PRM performed the experiments and molecular analyses, and conducted the statistical analyses. ARB, MQ and JH participated in the conceptual development and discussion of the study. YML, EL and AL wrote the manuscript with input from all authors.

Table 1: Summary statistics for the C/EBP-1 gene, overall and per region along the latitudinal gradient

region	n	A	pA	A _R	H _o	F _{is}	Theta K	S	π
Lower Saxony	86	23 (6)	11	21.5	0.826	0.08	2.62	12	0.003
Skåne	75	16 (5)	7	14.6	0.760	-0.06	2.48	9	0.003
Uppland	73	14 (4)	5	13.1	0.616	0.20	3.30	9	0.004
Västerbotten	68	3 (2)	0	2.9	0.574	-0.16	1.00	1	0.001
Norrbottn	62	4 (3)	2	4	0.065	-0.02	2.00	4	0.003
Total	364	38 (11)			0.593	0.04	2.70	16	0.004

n is the number of genotyped individuals, A indicates the number of nucleotide alleles with the number of unique amino acid sequences in brackets, pA is the number of private nucleotide alleles, A_R is the allelic richness corrected for the smallest samples size of 62, H_o stands for observed heterozygosity. F_{is} is the inbreeding coefficient, Theta K is the average number of pairwise nucleotide differences among alleles, S stands for the number of segregating sites and π is the nucleotide diversity.

Table 2: **Pairwise G'_{ST}** values among the five study regions with 95% confidence intervals, based on 1000 random neutral SNP markers (above the diagonal) and the C/EBP-1 allele frequencies (below diagonal). For all pairwise comparisons, the G'_{ST} value for the C/EBP-1 gene is significantly higher than for the neutral SNPs, with one exception (Skåne-Uppland) marked in blue.

SNP C-EBP	South			North		South n=81	North n=55
	L. Saxony n=30	Skåne n=26	Uppland n=25	Västerbotten n=28	Norrbottn n=27		
Lower Saxony n=86		0.174 (0.164-0.186)	0.185 (0.174-0.198)	0.291 (0.282-0.303)	0.375 (0.364-0.388)		
Skåne n=75	0.341 (0.213-0.480)		0.150 (0.140-0.161)	0.317 (0.307-0.329)	0.422 (0.411-0.438)		
Uppland n=73	0.356 (0.237-0.491)	0.127 (0.048-0.227)		0.282 (0.273-0.293)	0.403 (0.391-0.416)		
Västerbotten n=68	0.874 (0.810-0.929)	0.979 (0.954-0.995)	0.821 (0.735-0.891)		0.160 (0.149-0.173)		
Norrbottn n=62	0.970 (0.939-0.993)	0.997 (0.989-1.00)	0.997 (0.989-1.00)	0.379 (0.268-0.501)			
South n=234							0.267 (0.253-0.285)
North n=130						0.945 (0.920-0.967)	

Table 3. Signatures of historical selection on the C/EBP-1 gene measured as the dN/dS ratio (ω) and Tajima's D.

	A	dN	dS	ω	Z	p	Tajima's D	p
L. Saxony	23	0.0018	0.0157	0.115*	-2.60	0.011	-0.893	> 0.1
Skåne	16	0.0007	0.0131	0.053*	-2.41	0.017	-0.667	> 0.1
Uppland	14	0.0011	0.0170	0.065*	-2.44	0.017	0.633	> 0.1
Västerbotten	3	0.0013	0.0021	0.619	-0.97	0.330	n/a	n/a
Norrbottn	4	0.0019	0.0062	0.311	-0.98	0.327	-0.780	> 0.1
total	38	0.0019	0.0159	0.119*	-2.60	0.010	-1.092	> 0.1

A indicates the number of alleles, dN and dS are the relative rates of non-synonymous and synonymous substitutions, respective. ω is the ratio of dN/dS, Z is the test statistic, the first p gives the probability that Z is observed under neutral expectations, with * indicating significant deviation from neutrality. The second p refers to Tajima's D; this could not be calculated for the Västerbotten region, as a minimum of four sequences is required for the calculation.

Table 4. ANOVA models explaining variation at two phenotypic traits, larval period and mass at metamorphosis, with genetic variation at the C/EBP-1 gene.

predictor	Larval period				Mass			
	F _(1,111)	p-value	estimate	std error	F _(1,111)	p-value	estimate	std error
MCA 1	66.12	<0.0001	4.566	0.561	101.77	<0.0001	-0.0898	0.0089
MCA 2	24.84	<0.0001	-3.183	0.638	2.49	0.117	0.0159	0.0101

Based on 113 moor frog families from all regions along the gradient. Predictors for genetic variation are MCA1 and MCA2, the first two axes of the multiple correspondence analysis illustrated in Figure 5. Phenotypic traits were measured in a common garden experiment. Given are the F-value, the p-value, the estimate and the corresponding standard error for each factor. Significant values are given in bold.

Table 5. Linear regression results from the NOIA model (Alvarez-Castro & Carlborg 2007) investigating the association between the C/EBP-1 genotype and the larval phenotype for a) the northern regions of the gradient (Västerbotten and Norrbotten) and b) the southern regions of the gradient (Lower Saxony, Skåne and Uppland).

a)

allele		Larval period			Mass at metamorphosis			Freq [%]
		Effect	Std error	p-value	Effect	Std error	p-value	
A01	a	-2.146	0.531	0.0002	-0.026	0.010	0.0121	75.0
A04	a	2.178	0.571	0.0004	0.029	0.010	0.0068	23.0

b)

allele		Larval period			Mass at metamorphosis			Freq [%]
		Effect	Std error	p-value	Effect	Std error	p-value	
A02	a							40.7
A03	a							23.1
A04	a							15.7
A06	a	-2.712	0.965	0.0067				7.4

Based on 54 (a) and 59 (b) families with genotypic and phenotypic data available, and for alleles occurring in >3% frequency. Genetic effects with p-values <0.05 are given, all effects significant after Bonferroni correction. a stands for the additive genetic effect.

Figure 1. Frequencies of the identified C/EBP-1 alleles at the 15 collection sites along the latitudinal gradient, based on 364 genotyped *R. arvalis* individuals. For better visibility, colours have been applied only to the 13 C/EBP-1 alleles shared among at least two regions. Black arrows symbolize the two post-glacial recolonizations route into Scandinavia.

Figure 2. Network of the 38 C/EBP-1 alleles based from 364 genotyped individuals. The size of the circles corresponds to the relative frequencies of the alleles, and colours indicate their occurrence in the different geographical regions. Major alleles are marked with names, each tick mark indicates a mutational step.

Figure 3. Analysis of selection along 234 codons (702 bp) of the C/EBP-1 gene, calculating dN/dS (ω) for each codon, based on different assumptions for a, b and c. (a) using OmegaMap (Wilson & McVean 2006), (b) using CodeML (Yang 2007) without defining a phylogenetic tree, and (c) using CodeML (Yang 2007) with the best phylogenetic tree. The mean estimate of ω (=dN/dS) is indicated by the black line, and the grey shading in (a) represents the 95% highest posterior probability density (HPD). The positions marked with stars are under significant positive selection.

Figure 4. Phenotypic variation along the latitudinal gradient. Given are the mean values per region with standard error.

Figure 5. Multiple correspondence analysis for the 113 *R. arvalis* families based on the presence or absence of the five C/EBP-1 alleles occurring in more than 3% frequency. MCA1 on the x-axis explains 23.0%, and MCA2 on the y-axis explains 17% of the variation. Figure (A) shows how individuals genetically group according to their geographic origin, and (B) sorts the different genotypes according to MCA1 and MCA2.