Evolutionary adaptation to climate in microtine mammals

Remco Folkertsma

Univ.-Diss.

zur Erlangung des akademischen Grades
“doctor rerum naturalium”
(Dr. rer. nat)
in der Wissenschaftsdisziplin “Evolutionsgenetik”

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
Institut für Biochemie und Biologie
der Universität Potsdam

Datum der Disputation: 31. August 2020

Haubtbetreuer:	Prof. Dr. Michael Hofreiter (University of Potsdam)
Zweitbetreuer:	Prof. Dr. Jana A. Eccard (University of Potsdam)
Weitere Gutachter:	Prof. Dr. Jeremy Searle (Cornell University)
	Prof. Dr. Michael T. Monaghan (Leibniz IGB)

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-47680
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-476807

Summary

Understanding how organisms adapt to their local environment is a major focus of evolutionary biology. Local adaptation occurs when the forces of divergent natural selection are strong enough compared to the action of other evolutionary forces. An improved understanding of the genetic basis of local adaptation can inform about the evolutionary processes in populations and is of major importance because of its relevance to altered selection pressures due to climate change. So far, most insights have been gained by studying model organisms, but our understanding about the genetic basis of local adaptation in wild populations of species with little genomic resources is still limited.

With the work presented in this thesis I therefore set out to provide insights into the genetic basis of local adaptation in populations of two voles species: the common vole (*Microtus arvalis)* and the bank vole (*Myodes glareolus)*. Both voles species are small mammals, they have a high evolutionary potential compared to their dispersal capabilities and are thus likely to show genetic responses to local conditions, moreover, they have a wide distribution in which they experience a broad range of different environmental conditions, this makes them an ideal species to study local adaptation.

The first study focused on producing a novel mitochondrial genome to facilitate further research in *M. arvalis*. To this end, I generated the first mitochondrial genome of *M. arvalis* using shotgun sequencing and an iterative mapping approach. This was subsequently used in a phylogenetic analysis that produced novel insights into the phylogenetic relationships of the Arvicolinae.

The following two studies then focused on the genetic basis of local adaptation using ddRAD-sequencing data and genome scan methods. The first of these involved sequencing the genomic DNA of individuals from three low-altitude and three high-altitude *M. arvalis* study sites in the Swiss Alps. High-altitude environments with their low temperatures and low levels of oxygen (hypoxia) pose considerable challenges for small mammals. With their small body size and proportional large body surface they have to sustain high rates of aerobic metabolism to support thermogenesis and locomotion, which can be restricted with only limited levels of oxygen available. To generate insights into high-altitude adaptation I identified a large number of single

nucleotide polymorphisms (SNPs). These data were first used to identify high levels of differentiation between study sites and a clear pattern of population structure, in line with a signal of isolation by distance. Using genome scan methods, I then identified signals of selection associated with differences in altitude in genes with functions related to oxygen transport into tissue and genes related to aerobic metabolic pathways. This indicates that hypoxia is an important selection pressure driving local adaptation at high altitude in *M. arvalis*. A number of these genes were linked with high-altitude adaptation in other species before, which lead to the suggestion that high-altitude populations of several species have evolved in a similar manner as a response to the unique conditions at high altitude

The next study also involved the genetic basis of local adaptation, here I provided insights into climate-related adaptation in *M. glareolus* across its European distribution. Climate is an important environmental factor affecting the physiology of all organisms. In this study I identified a large number of SNPs in individuals from twelve *M. glareolus* populations distributed across Europe. I used these, to first establish that populations are highly differentiated and found a strong pattern of population structure with signal of isolation by distance. I then employed genome scan methods to identify candidate loci showing signals of selection associated with climate, with a particular emphasis on polygenic loci. A multivariate analysis was used to determine that temperature was the most important climate variable responsible for adaptive genetic variation among all variables tested. By using novel methods and genome annotation of related species I identified the function of genes of candidate loci. This showed that genes under selection have functions related to energy homeostasis and immune processes. Suggesting that *M. glareolus* populations have evolved in response to local temperature and specific local pathogenic selection pressures.

The studies presented in this thesis provide evidence for the genetic basis of local adaptation in two vole species across different environmental gradients, suggesting that the identified genes are involved in local adaptation. This demonstrates that with the help of novel methods the study of wild populations, which often have little genomic resources available, can provide unique insights into evolutionary processes.

Zusammenfassung (German summary)

Ein Schwerpunkt der Evolutionsbiologie besteht darin, zu verstehen, wie sich Organismen an ihre lokale Umgebung anpassen. Lokale Anpassung tritt ein, wenn die Kräfte der divergierenden natürlichen Selektion im Vergleich zu anderen evolutionären Kräften stark genug sind. Ein verbessertes Verständnis der genetischen Grundlagen der lokalen Anpassung kann Informationen über die Evolutionsprozesse in Populationen liefern und ist durch seine Relevanz für durch den Klimawandel bedingte veränderte Selektionsdrücke von großer Bedeutung. Bisher wurden die meisten Erkenntnisse durch Untersuchungen an Modellorganismen gewonnen. Jedoch ist das Verständnis der genetischen Grundlagen der lokalen Anpassung in Wildpopulationen von Arten mit geringen genomischen Ressourcen noch immer begrenzt.

Mit den in dieser Doktorarbeit vorgestellten Untersuchungen war es daher mein Ziel, Einblicke in die genetischen Grundlagen der lokalen Anpassung in Populationen von zwei Wühlmausarten zu geben: der Feldmaus (*Microtus arvalis*) und der Rötelmaus (*Myodes glareolus*). Bei beiden handelt es sich um kleine Säugetiere mit einem, im Vergleich zu ihrer Ausbreitungsfähigkeit, hohen Evolutionspotential. Daher ist anzunehmen, dass sie genetische Reaktionen auf lokale Bedingungen zeigen. Hinzu kommt, dass sie aufgrund ihrer großen Verbreitung ein großes Spektrum an verschiedenen Umweltbedingungen erfahren, was sie zu einer idealen Spezies, für die Untersuchung lokaler Anpassung macht.

Die erste Studie dieser Arbeit konzentrierte sich auf die Erstellung eines bisher nicht verfügbaren mitochondriellen Genoms, um die weitere Forschung an *M. arvalis* zu erleichtern. Dies wurde mittels Shotgun-Sequenzierung und eines iterativen Kartierungsansatzes erreicht. Anschließend wurde es in einer phylogenetischen Analyse verwendet, die neue Erkenntnisse über die phylogenetischen Beziehungen der Arvicolinae lieferte.

Die folgenden zwei Studien konzentrierten sich auf die genetische Basis der lokalen Anpassung unter Verwendung von ddRAD-Sequenzierungsdaten und Genom-Scan-Methoden. Die erste umfasste die Sequenzierung der genomischen DNA von Individuen aus drei *M. arvalis*-Untersuchungsgebieten in geringer Höhe und drei in großer Höhe in den Schweizer Alpen. Umgebungen in großer Höhe mit niedrigen Temperaturen und niedrigem Sauerstoffgehalt

(Hypoxie) stellen kleine Säugetiere vor erhebliche Herausforderungen. Aufgrund ihrer geringen Körpergröße und proportional großen Körperoberfläche müssen sie hohe aerobe Stoffwechselraten aufrechterhalten, um die Thermogenese und Fortbewegung zu unterstützen, die mit begrenzter Sauerstoffverfügbarkeit eingeschränkt sein können. Um Einblicke in die Höhenanpassung zu erhalten, habe ich eine große Anzahl von Einzelnukleotidpolymorphismen (SNPs) identifiziert. Mit Hilfe dieser Daten wurden ein hohes Maß an Differenzierung zwischen den Untersuchungsorten und ein klares Muster der Populationsstruktur zusammen mit einem isolation-by-distance Signal identifiziert. Unter Verwendung von Genom-Scan-Methoden identifizierte ich Selektionssignale in Genen, die mit Höhenunterschieden verbunden werden. Diese besitzen Funktionen, die mit dem Sauerstofftransport in das Gewebe sowie mit aeroben Stoffwechselwegen zusammenhängen. Dies weist darauf hin, dass Hypoxie ein wichtiger Selektionsdruck für die lokale Anpassung in großer Höhe für *M. arvalis* ist. Einige dieser Gene sind bereits früher mit der Höhenanpassung bei anderen Arten in Verbindung gebracht worden. Dies führte zu der Annahme, dass sich Populationen in großer Höhe lebender verschiedener Arten in Anpassung an die einzigartigen Bedingungen in großer Höhe auf ähnliche Weise entwickelt haben.

Die nächste Studie befasste sich ebenfalls mit den genetischen Grundlagen der lokalen Anpassung. Hier stellte ich Erkenntnisse über die klimabedingte Anpassung von *M. glareolus* in ihrem europäischen Verbreitungsgebiet vor. Das Klima ist ein wichtiger Umweltfaktor, der die Physiologie aller Organismen beeinflusst. In dieser Studie identifizierte ich zehntausende SNPs bei Individuen aus zwölf in ganz Europa verteilten *M. glareolus*-Populationen. Diese ergaben eine starke Differenzierung der Populationen mit deutlicher Populationsstruktur und einem Signal für isolation-by-distance. Anschließend verwendete ich Genom-Scan-Methoden, um mögliche Loci zu identifizieren, die mit dem Klima verbundene Selektionssignale aufweisen, wobei der Schwerpunkt dabei auf polygenen Loci lag. Eine Multivariaten Analysemethode ermittelte, dass die Temperatur die wichtigste Klimavariable unter allen getesteten Variablen ist, die für die adaptive genetische Variation verantwortlich ist. Mit Hilfe neuartiger Methoden und der Annotation von Genomen verwandter Spezies identifizierte ich die Funktion von Genen an Kandidatenloci. Diese zeigten, dass die unter Selektion stehenden Gene Funktionen im Zusammenhang mit der Energiehomöostase und den Immunprozessen ausüben. Dies wiederum deutet darauf hin, dass sich die Populationen von *M. glareolus* in

Reaktion auf die lokale Temperatur und den spezifischen lokalen Selektionsdruck für Krankheitserreger entwickelt haben.

Die in dieser Arbeit vorgestellten Studien liefern Belege für die genetische Basis der lokalen Anpassung auf verschiedene Umweltgradienten in zwei Wühlmausarten. Dies deutet darauf hin, dass die identifizierten Gene an der lokalen Anpassung beteiligt sind. Darüber hinaus zeigt dies, dass Untersuchungen wildlebender Populationen mit geringen genomischen Ressourcen durch den Einsatz neuartiger Methoden einzigartige Einblicke in evolutionäre Prozesse ermöglichen können.

Acknowledgements

First of all, I would like to thank my supervisors Prof. Dr. Michael Hofreiter and Prof. Dr. Jana Eccard for giving me the opportunity to come to Germany and to work on such an interesting topic. Without your guidance and support, I would not be here today.

Furthermore, I would like to thank ….
…. My collaborators on the different projects in this thesis, specifically: Gerald Heckel, Petr Kotlík and Nathalie Charbonnel.
…. Prof. Dr. Michael Lenhard, for trusting me to work in his lab and running qPCRs without dye.

…. The Evolutionary Adaptive Genomics group, present and past members,
…. George, for being an awesome office-mate and my best Greek friend,
…. Nik, for his role as a coffee-partner,
…. Johanna, my fellow Dutchie, for being my lively subgroup-leader,
…. Axel, for his couch and his help with sequence alignments,
…. Stefanie, for introducing me to bioinformatics and her “bless you’s” from across the hall,
…. The people on the balcony, thank you for all the moments of relaxation
…. the people who took time out of their busy lives to proof-read my thesis: Elis, Fede and Ulrike.

…. All members of the Animal Ecology group in the Maulbeerallee, thank you for offering me a nice place to work outside of Golm.

Further, I would like to thank all members of the bio-group: Christiaan, Maya, Arne, Hanneke, Michiel, Inge, Patrick, Aline, and Floris for coming over and having me whenever I came back to the Netherlands.

A special thanks to my mum, dad and sister for encouraging me to go to Potsdam and providing me with the support to pursue my dreams.

Last but not least I want to thank Anna Walther. You helped me through all the rough times during my PhD. Thank you for your optimism and cheerfulness, without you I couldn’t have done it, I wouldn’t change a thing.

Contents

Abbreviations

AFLP	Amplified fragment length polymorphism
BMR	Basal metabolic rate
F_{ST}	Fixation index
GEA	Genotype-environment association
IBD	Isolation by distance
LD	Linkage disequilibrium
LGM	Last glacial maximum
mtDNA	Mitochondrial DNA
MYA	Million years ago
NGS	Next generation sequencing
NST	Non-shivering thermogenesis
PCA	Principal component analysis
RDA	Redundancy analysis
SNP	Single nucleotide polymorphism

Chapter 1: Introduction

1.1 Local adaptation

Uncovering the genetic basis of local adaptation is a major focus of evolutionary biology. Local adaptation occurs when individuals from a species have higher fitness in their native environment than any individuals from the same species introduced from elsewhere (Kawecki & Ebert 2004; Blanquart et al. 2013; Savolainen et al. 2013; Tiffin & Ross-Ibarra 2014). The driving force of local adaptation originates from spatially and temporally varying natural selection. Natural selection acts on the phenotypes of individuals and can lead to genetic differences between populations (Richardson et al. 2014). Over time, local adaptation can maintain genetic variation, aid species range expansion and eventually results in ecological speciation (Hereford 2009; Savolainen et al. 2013). An understanding of the genetic basis of local adaptation is important to understand evolution and the processes that have shaped adaptive responses, as well as to how species will respond to changing environmental conditions (Gienapp et al. 2008; Hoffmann & Willi 2008; Hoffmann & Sgrò 2011; Franks & Hoffmann 2012; Pauls et al. 2013).

The capacity for species and populations to locally adapt is dependent on their evolutionary potential, which is dependent on species' generation time and effective population size. Generation time is an important factor, as it determines the speed at which populations can evolve in response to changes in their environment (Réale et al. 2003; Visser 2008). In general, small species with short generation times will be better able to genetically track changes in their environment than species with longer generation times. Natural selection will only cause evolution if there is enough genetic variation, which can be caused by mutations, recombination and gene flow (Gandon & Michalakis 2002; Barrett & Schluter 2008; Hoffmann & Sgrò 2011). The amount of standing genetic variation is highly dependent on effective population size, because the effects of drift are less severe in larger populations and the likelihood to acquire new mutations is higher (Hedrick 2011). Accordingly, in populations that go through frequent population crashes and bottlenecks, genetic drift may lower the amount of genetic variation, hampering the evolution of local adaptation (Kawecki & Ebert 2004). Local adaptation is more difficult in populations that are exposed to high levels of gene flow (Lenormand 2002; Bourne et al. 2014), therefore, species with low dispersal capabilities are more likely to adapt to local conditions. Thus, small species, with low dispersal capabilities, short generation times, large

effective population sizes and therefore, large evolutionary potential, are expected to have a high potential to locally adapt.

The evolution of local adaptation requires that alleles which increase fitness in one environment, should decrease fitness in another environment. If this were not the case, the allele with the highest fitness would invade all populations and eventually become monomorphic. Local adaptation is facilitated by spatially varying selection across the landscape, but also influenced by genetic drift and gene flow (Kawecki & Ebert 2004; Hereford 2009; Hedrick 2011). Genetic drift can hinder local adaptation by removing genetic variation or fixating random alleles and thereby produce random variation that confounds local adaptation (Lande 1976; Travisano et al. 1995). Gene flow from neighboring populations can introduce new (and possibly maladaptive) alleles that may swamp locally favorable alleles (Lenormand 2002). However, gene flow also has the potential to aid local adaptation by introducing new genetic variation for selection to act upon and alleviate the confounding effects of genetic drift on local adaptation (Garant et al. 2007; Savolainen et al. 2007). Thus, if the strength of selection is large enough and the amount of gene flow from non-adapted populations is not too high, local adaptation should occur.

The presence of local adaptation can be tested with field experiments. Reciprocal transplant experiments have provided evidence that populations are locally adapted and common garden experiments have shown that many of the phenotypic differences between populations are heritable (Berven 1982; Hereford 2009). These approaches, however, do not provide insight into the genetic basis of local adaptation. Some of the earliest studies shedding light on the genes or genetic regions responsible for local adaptation did so by focusing on traits that were already thought to be targets of selection. These either focused on a clear phenotypic trait, such as the change in color variation during the industrial revolution in peppered moth (*Biston betularia*) (Cook & Saccheri 2013; Hof et al. 2016), and the adaptive variation in coat color in pocket mice (*Chaetodipus intermedius*) (Nachman et al. 2003) and deer mice (*Peromyscus maniculatus*) (Linnen et al. 2009). Or by focusing on candidate genes, such as those responsible for adaptation to high altitude in deer mice (Storz et al. 2007) and flowering time in *Arabidopsis thaliana* (Le Corre & Kremer 2012).

The classical methods applied to detect local adaptation use a top-down approach, relying on fitness measurements or phenotypic variation and relate this to genotypic data. Quantitative trait

locus mapping approaches often study the progeny of crosses between different locally adapted populations and correlate the variation in a quantitative trait in the phenotype with genotypic variation (Mackay 2001; Bernardo 2008). Genome wide association studies evaluate the association between genetic markers and a phenotype of interest, or a measurement of fitness, within a large number of individuals, in order to identify genetic markers that are associated with these traits in the field (Korte & Farlow 2013). Such methods and experiments require space, time, and genomic resources that are not available for many species. Therefore, the use of these methods has mainly been restricted to model organisms (Blanquart et al. 2013). Thus, although we understand the genetic basis for local adaptation in model organisms, the molecular basis of local adaptation in non-model organisms is still poorly understood. However, recently the approaches that have been used for decades have changed, as cost-effective methods that allow for the production of large amounts of genetic data have been developed. With these in place, it is now more feasible for researchers to conduct studies on local adaptation also with genetic data rather than with phenotypic data alone.

1.2 Next generation sequencing

The advent of next generation sequencing (NGS)-technologies has led to an ever-increasing availability of genomic resources for non-model organisms. NGS opened up the possibility to sequence millions of DNA fragments in parallel, greatly increasing the speed at which genotypic data can be generated for a relatively low price. This allows for the screening of thousands to millions of single nucleotide polymorphisms (SNPs) across the entire genome in non-model organisms and wild populations at reasonable costs.

In the pre-NGS era, researchers studying non-model organisms often used the amplified fragment length polymorphism (AFLP)-technique to obtain genetic markers and scan the genome for signals of selection, because this approach does not require prior information about the genome. Although generating AFLP-markers is relatively easy, linking markers of interest to genes under selection and generating enough markers to span the entire genome is difficult using this approach. Therefore, while studies using AFLP-markers have identified signals of selection, they have largely failed to identify the genes or genomic regions underlying local adaptation (Bonin 2008), unless additional sequencing efforts were made (e.g. Buehler, Holderegger, Brodbeck, Schnyder, & Gugerli, 2014; Wood, Grahame, Humphray, Rogers, & Butlin, 2008). However, these limitations disappeared with the advent of NGS and the

accompanying rapid development of genomic tools, resulting in numerous studies using SNPs as genetic markers to study local adaptation in non-model organisms (Haasl & Payseur 2016).

1.3 Genome scan methods

Since genomic data has become readily available, researchers have begun to study local adaptation starting from genetic data (i.e. bottom-up approaches), by using population and landscape genomic approaches. These approaches do not require prior knowledge of genes or loci subject to local adaptation, and can thus be used without the need for a reference genome (although this does aid subsequent analyses). The field of landscape genomics has developed a number of tools to test markers across the genome for signatures of selection and study local adaptation (Rellstab et al. 2015; Hoban et al. 2016). These genome-scan methods can broadly be defined in two categories: Differential outlier (F_{ST})-methods and genotype environmental association (GEA)-methods.

If there is no prior information on local conditions or traits driving adaptation, F_{ST}-based methods can be used to detect loci under positive selection. Alleles favoring local adaptation should occur at higher frequency where they are beneficial and thus increase fitness, while they should occur at lower frequency where they are detrimental and thus decrease fitness. Therefore, loci involved in local adaptation may be identified by screening the genome for loci showing above average differentiation among populations, usually measured as F_{ST}. Loci with extreme F_{ST}-values are consequently suggested to be under positive selection (Lewontin & Krakauer 1973). In contrast, GEA-methods can be applied when the environmental variables of interest leading to local adaptation are known (or at least hypothesized). These methods search for alleles among populations with allele frequencies that are strongly associated with environmental conditions. Such a correlation between allele frequencies and environment suggests that these loci are involved in local adaptation (Mitton et al. 1977).

Many studies have provided insight into adaptive genetic variation using F_{ST} and GEA-based methods. However, despite advances in the development of genome scan methods, detecting adaptive genetic variation remains challenging with both approaches (Harrisson et al. 2014). Theoretical and empirical studies suggest that the traits responsible for adaptation to climate are commonly of polygenic nature and are underpinned by many genes of small effect (Pritchard & Di Rienzo 2010; Yeaman 2015; Wellenreuther & Hansson 2016; Babin et al. 2017;

Harrisson et al. 2017). However, genome scans used to detect adaptive genetic variation generally lack the power to detect weak signals of polygenic selection (Rellstab et al. 2015; Wellenreuther & Hansson 2016; Forester et al. 2018). This is because genome scans only focus on one locus at a time and not on the combined signal of all loci simultaneously. Recently, multivariate methods from the field of community ecology have been introduced to the field of population genomics. These methods are able to focus on many loci and multiple environmental predictors at the same time and are thus better suited to detect signals of polygenic selection (Forester et al. 2018; Capblancq et al. 2018).

So far, numerous tests have been developed to identify loci under positive selection (e.g. Beaumont & Balding, 2004; de Villemereuil & Gaggiotti, 2015; Foll & Gaggiotti, 2008; Luu, Bazin, & Blum, 2017; Whitlock & Lotterhos, 2015) or to identify loci with allele frequencies correlating with environmental traits (e.g. Frichot & François, 2015; Frichot, Schoville, Bouchard, & François, 2013; Gautier, 2015; Joost et al., 2007). Many of these methods are widely used, but it has to be taken into account that all have their own advantages and disadvantages (de Villemereuil et al. 2014; Lotterhos & Whitlock 2015) and use different approaches to control for confounding effects.

Genome scan methods have to distinguish loci under selection from loci showing genetic differentiation between populations caused by neutral forces. This can be difficult as demographic history and neutral processes such as gene flow and genetic drift can produce allele frequency differences between populations that look similar to those caused by positive selection or local adaptation. These confounding patterns of neutral population structure can mimic patterns of local adaptation and have to be corrected for, in order to prevent incorrect conclusions. This can be particularly challenging in species with low dispersal capabilities and low gene flow between populations, as these species are often highly structured and the average level of between population differentiation is high. Because this results in high F_{ST}-values for loci that are selectively neutral, loci under selection need to have extremely high F_{ST}-values to be considered as outlier under selection.

Another confounding effect of demographic history that may cause spurious association between allele frequency and variables of interest, is caused by the colonization of new areas, which results in a phenomenon called allele-surfing (Edmonds et al. 2004; Klopfstein et al. 2006). Allele-surfing occurs when populations expand and populations at the front wave of the

expansion experience strong genetic drift. This causes a decrease in genetic diversity along the axis of expansion and clinal patterns of differentiation in allele frequency over the landscape that resemble patterns of positive selection (Excoffier & Ray 2008; Excoffier et al. 2009). Furthermore, hybridization and introgression from related taxa can also cause allele frequency patterns that confound detecting of local adaptation (Kane et al. 2009; Geraldes et al. 2014).

Finally, confounding effects can arise from various types of autocorrelation that stem from the shared history of sampled populations. For example, there is often spatial autocorrelation of environmental variables. Nearby environments are usually more similar than environments further apart and populations are often more related to each other over close than over long distances. This can lead to associations between neutral alleles and environmental variables, that are caused by the mere spatial arrangement or demographic history of populations, but are not signatures of local adaptation. A similar pattern can occur due to range expansion from glacial refugia (Hewitt 1999). This can create spatial patterns in allele frequencies that coincide with environmental variables such as temperature, that cause spurious correlations between allele frequencies and environmental variables of interest (Holliday et al. 2010). Lastly, background selection, the process in which neutral alleles linked with deleterious alleles are removed with purifying selection (Charlesworth et al. 1993), may cause stronger differentiation between populations at linked neutral sites, that can be mistakenly recognized as patterns of positive selection favoring local adaptation (Charlesworth et al. 1997; Cruickshank & Hahn 2014).

Several approaches have been developed to address the problem of confounding effects of neutral processes and spatial autocorrelation. These either make use of explicit models to infer demographic history or use the relationship of the sampled populations to correct for neutral genetic structure. The first type of approaches simulates a null distribution based on the island model (Beaumont & Balding 2004) or a multinomial Dirichlet distribution (Foll & Gaggiotti 2008). This distribution is then tested against the observed data to infer loci influenced by local adaptation. However, if the assumptions of these models are not met, they can suffer from high false positive rates (De Mita et al. 2013; de Villemereuil et al. 2014; Lotterhos & Whitlock 2014). More recent approaches make use of the observed genetic data to explicitly correct for relationships among populations, by estimating covariance or relatedness among them. Neutral processes like demography and drift affect all loci across the genome, whereas positive

selection and other non-neutral processes only affect a subset of loci. Several methods make use of this disparity to estimate neutral structure. Ideally, a presumed set of genetic neutral markers is known, but since this is usually not the case, the observed genetic data is used. These analyses can be performed in multiple ways, for example, by using a pairwise covariance matrix (Coop et al. 2010; Günther & Coop 2013; Gautier 2015), a population tree (Bonhomme et al. 2010), or using linear models (Frichot et al. 2013; Frichot & François 2015; Luu et al. 2017). Approaches using a covariance matrix or population tree use these to generate a null distribution and to identify loci that depart from this. Linear model approaches incorporate neutral genetic structure by including principal components (Luu et al. 2017), random effects or so-called latent factors (Frichot & François 2015). Approaches that infer confounding effects from the observed data are more flexible as they do not assume an explicit demographic model, and in general perform better than models based on demographic null models (de Villemereuil et al. 2014; Lotterhos & Whitlock 2014). Altogether genome-wide datasets in combination with genome scans provide a powerful approach to study adaptation in heterogeneous environments.

1.4 Relevant environmental variation

Spatial environmental variation is omnipresent and populations are usually adapted to local conditions (Savolainen et al. 2013). Evidence for selection acting across diverse environmental conditions has been found in a wide range of taxa (Franks & Hoffmann 2012). Environmental clines offer an ideal system to study the genetic underpinnings of local adaptation, as changes in selection pressure along a cline result in locally adapted populations (Slatkin 1973; Bridle et al. 2010). In this thesis, I therefore focus on the genetic basis of local adaptation across two of these environmental gradients: A latitudinal gradient in climate across Europe and an altitudinal gradient in the Swiss Alps.

1.4.1 Climate and temperature

Climate is a key environmental factor affecting the physiology and metabolism of all species involved. Because of its widespread effect, recent climate change has become a focus of evolutionary ecology (Hoffmann & Sgrò 2011; Franks & Hoffmann 2012; Merilä 2012). Phenotypic changes in response to climate change have been found in a range of taxa, but overall it seems that these are a result of phenotypic plasticity, rather than the result of genetically based changes (Merilä & Hendry, 2014 and accompanying reviews). In contrast, there is widespread evidence for local adaptation and adaptive clines in response to spatial

variation in climatic conditions associated with latitudinal and altitudinal gradients (Conover et al. 2009). Within species, phenotypes tend to change in a somewhat predictable way along climatic gradients. Such changes include latitudinal clines in body size (Bergmann 1848; Huey et al. 2000) and the phenology of many plant species (Olsson & Ågren 2002; Stinchcombe et al. 2004). There is evidence for genetic adaptation to heat tolerance across thermal gradients associated with latitude (Hoffmann et al. 2002; De Jong et al. 2013; Diamond et al. 2017). A multitude of studies have identified genes and pathways that are potentially involved in climate adaptation, these include changes in genes related to thermal tolerance, drought resistance and other traits related to climate such as pathogen resistance (Franks & Hoffmann 2012). Although numerous studies have aided our understanding of adaptation to climate, the genetic basis for climate adaptation is complex, and studies in vertebrates with a widespread distribution are necessary for a broader understanding of the evolutionary basis for climate-adaptation.

1.4.2 Altitude

Local conditions can vary dramatically over much short spatial distances along altitudinal gradients than over latitudinal gradients, which offers an excellent opportunity to study local adaptation in the face of gene flow and without the confounding effects of demographic history (Keller et al. 2013). The most obvious changes along altitudinal gradients are a decrease in temperature, a reduction in oxygen availability and an increase in solar radiation intensity with increasing elevation. More specifically, ambient temperatures decrease by 6.5°C and the available oxygens levels drop by 10% roughly every 1,000 meters in elevation (Körner 2007). The harsh conditions at high altitude and rapid changes in conditions over short distances have fueled research using altitudinal gradients. Evidence for physiological adaptation to high altitude are widespread and involve changes in the pulmonary and respiratory system, oxygen-hemoglobin affinity and tissue capillarity (Storz et al. 2010). A genetic basis for high-altitude adaptation is determined in humans as well as in other vertebrate species (Storz et al. 2012; Simonson 2015; Wei et al. 2016; Bigham 2016; Song et al. 2016).

The reduction in oxygen availability at high altitude threatens aerobic metabolism and poses a major challenge that potentially limits the amount of oxygen available for activity (West 1990). For small endothermic mammals and birds this is particularly challenging, as they have to retain a constant body temperature in these cold conditions. At the same time, they only have a small tissue volume to produce heat compared to their large surface area via which they lose heat.

Therefore, small mammals are constantly challenged by high thermogenic demands. This has captured the interest of researchers and there is a large body of literature on the specific physiological and evolutionary adaptations of high-altitude small mammals (Storz et al. 2010; Cheviron & Brumfield 2012). A large part of this is specifically focused on high-altitude deer mice, as they have a broad elevational distribution from the coast up to ~ 4,300 meters above sea level. A number of studies revealed that populations of deer mice residing at different altitudes have evolved differences in aerobic performance and thermogenic capacities that contribute to high-altitude adaptation (reviewed in McClelland & Scott, 2018; Storz, Cheviron, McClelland, & Scott, 2019). However, despite ongoing progress in identifying the genetic basis of high-altitude adaptation, insights are mainly restricted to a few selected species such as humans and deer mice for which large genomic resources are available, Therefore, a thorough investigation including non-model species is needed to further our understanding in the evolutionary basis of high-altitude adaptations.

1.5 ddRAD-sequencing

With the rise of NGS, the costs of sequencing whole genomes have been greatly reduced. However, sequencing hundreds of individuals of non-model organisms from natural populations still remains prohibitively expensive. Therefore, several alternative genomic approaches exist that allow researchers to sequence a higher number of samples for a given budget. Most researchers currently use reduced representation approaches (defined as any method that sequences less than the whole genome) to study individuals from wild populations. Most widely used are methods that apply restriction enzymes to reduce the complexity of the genome, but alternative approaches include methods that only target specific regions of the genome such as RNA-seq and exome capture.

Restriction site-associated DNA sequencing (RAD-seq) approaches use restriction enzymes to cut at specific motifs in the genome and use NGS to sequence the genomic regions adjacent to the restriction cut-sites (reviewed in Andrews, Good, & Miller, 2016; Puritz et al., 2014). The original RADseq (Miller et al. 2007; Baird et al. 2008) uses one restriction enzyme to digest genomic DNA, after which mechanical shearing is used to reduce fragments to the appropriate length for sequencing. Since then several RAD-seq variants have been developed (discussed in Andrews et al., 2016). Here, I focus on double digest restriction-site associated DNA (ddRAD) sequencing, as it is the method employed in this thesis. This method has been widely used in

non-model studies to investigate questions in an ecological and evolutionary context. It has been used to provide insights into population genomics, landscape genomics, phylogenetics, epigenetics and the genomics of local adaptation (Kjeldsen et al. 2016; Vargas et al. 2017; Cooper & Uy 2017; Dimond et al. 2017; Maas et al. 2018; Weigand et al. 2018; Hoey & Pinsky 2018; Homola et al. 2019).

Preparation of a ddRAD library starts with obtaining high-molecular-weight genomic DNA, as degraded DNA may result in poorer sequencing results (Graham et al. 2015). The ddRAD protocol itself then starts with digestion of genomic DNA using two restriction enzymes. Usually, this consists of a "common cuter", that cuts frequently and a "rare cutter" that cuts less frequently, as a result of the length of the enzyme recognition cut-site. For example, a rough estimate suggests that a 6-bp cutter will cut every 4^6=4,096 bp, while an 8-bp cutter will cut every 4^8=65,536 bp. A combination of enzymes can be chosen to fine-tune the number of loci expected. After digestion, Illumina adapters are ligated to the sticky ends created by the restriction enzymes. These adapters include sample specific barcodes, which allow for multiplexing (pooling of individuals) after ligation, thereby greatly reducing time and costs of subsequent steps. Once multiplexed, a size selection step is used to isolate fragments of a specific length for sequencing. This is another step that can be used to fine-tune the number of expected loci to be sequenced. When a reference genome or that of a closely related species is available, specific software can be used to determine the expected number of loci of a certain fragment length (Lepais & Weir 2014). Consistency of size selection across libraries is crucial to produce a comparable set of loci across samples. Inconsistent size selection results in different sets of loci selected among different libraries, which potentially leads to high levels of missing genotypes between them. Finally, the resulting size selected libraries are PCR amplified to create libraries for sequencing.

1.6 Research species

The two vole species that are investigated in the present work are the common vole (*Microtus arvalis*) and the bank vole (*Myodes glareolus*). Both species belong to the rodent subfamily Arvicolinae of the family Cricetidae. Arvicolinae is a highly diverse subfamily that includes approximately 150 species of voles, lemmings and muskrats (Nowak 1999). Vole is a common name for a large fraction of the species within the Arvicolinae (70-100 species, depending on the classification used), including species from several tribes (Galewski et al. 2006). Voles are

a diverse group of small-bodied mouse-like species that are distributed throughout the Northern hemisphere, mainly in temperate and boreal environments (Le Galliard et al. 2012). Where they occur, vole species are often the dominant small mammal species and are therefore of significant ecological importance. In general, voles live above ground, traveling through the vegetation, underneath the snow or use elaborate burrow systems. They are active throughout the year and do not go into hibernation. Their diet mainly consists of seeds and plants and the occasional insect. Voles are iteroparous, they further have a short life-span which usually includes only one reproductive season. During this reproductive season, females can have multiple litters and several generations can overlap and reproduce simultaneously, resulting in strong annual population dynamics (Boyce & Boyce 1988; Getz et al. 2007; Eccard & Herde 2013). Voles are also an example of species exhibiting strong multi-annual population density fluctuations (Tkadlec & Stenseth 2001; Lambin et al. 2006; Inchausti et al. 2009; Rikalainen et al. 2012; Gauffre et al. 2014). The severity of these fluctuations varies across their distribution, but can result in 10- to 100-fold changes in density (Stenseth 1999). Surprisingly, despite reoccurring bottlenecks, this does not cause loss of genetic variation, probably due to increased gene flow from neighboring populations in the high-density phase (Berthier et al. 2006; Rikalainen et al. 2012; Martínková et al. 2013; Gauffre et al. 2014). Dispersal is mainly male-biased, and males tend to both disperse more often and travel longer distances, while females are more philopatric and stay closer to their burrow (Aars et al. 1998; Gerlach & Musolf 2000; Schweizer et al. 2007; Gauffre et al. 2009; Gauffre et al. 2014). As they are small-bodied animals, voles can often only travel short distances. Recurring bottlenecks and low effective dispersal rates result in high levels of population differentiation even at small geographic scales (Redeker et al. 2006; Schweizer et al. 2007; Borkowska et al. 2010; Guivier et al. 2011). The above applies to both species studied in this work. Below, I separately describe each species and its characteristics used in this work, combined with implications for studying local adaptation.

1.6.1 The common vole (*Microtus arvalis*)

Figure 1.1: *Microtus arvalis*

The common vole (*M. arvalis*) (Pallas, 1778) (Figure 1.1) belongs to the genus *Microtus*. This genus is one of the most speciose mammalian genera that resulted from one of the fastest radiations within the last 2 million years (Jaarola et al. 2004; Fink et al. 2010). *Microtus* originated in South Asia, from where several colonization waves to Europe and North America occurred with subsequent radiations on each continent (Fink et al. 2010). *M. arvalis* has a continuous distribution across the European continent from the Atlantic coast of France to the steppes of Central Russian, and from the Baltic sea coast in the north to the Mediterranean coast in Spain in the south (Figure 1.2), where it is found from sea level to above 2,000 m a.s.l. (Fischer et al. 2011; Fischer et al. 2014), occupying dry habitats such as fields and meadows. Previous examinations using mitochondrial and nuclear DNA have revealed a genetic subdivision into several highly divergent evolutionary lineages in Europe (Haynes et al. 2003; Fink et al. 2004; Jaarola et al. 2004; Heckel et al. 2005; Hamilton et al. 2005; Tougard et al. 2008; Bužan et al. 2010; Martínková et al. 2013; Lischer et al. 2014; Stojak et al. 2016). In general, five evolutionary lineages have been identified: Western, Central, Eastern, Italian and Balkan, although some authors suggest an additional split of the Western lineage into a Western-North and Western-South lineage. The divergence among lineages likely occurred before the last glacial maximum (LGM), but the major driver is geographic isolation of surviving populations in multiple glacial refugia during the LGM (Lischer et al. 2014; García et al. 2019). Interestingly, studies showed that *M. arvalis* did not only retreat south during the LGM (Hewitt 1999), but also survived in locations further north (Fink et al. 2004; Hamilton et al. 2005; Pedreschi et al. 2019). Evidence suggests that the current spatial distribution of lineages is related to the climatic conditions to which the different populations are better adapted. However, whether these adaptations evolved before, during or after the LGM cannot be distinguished (Fink et al. 2004; Bize et al. 2018; Stojak et al. 2019). Recent studies have found loci potentially involved in adaptation to high altitude in alpine regions (Fischer et al. 2011; Fischer et al. 2014). However, as currently no

annotated reference genome exists for *M. arvalis*, neither putative functions of loci under selection nor the genetic basis of specific adaptations could be identified.

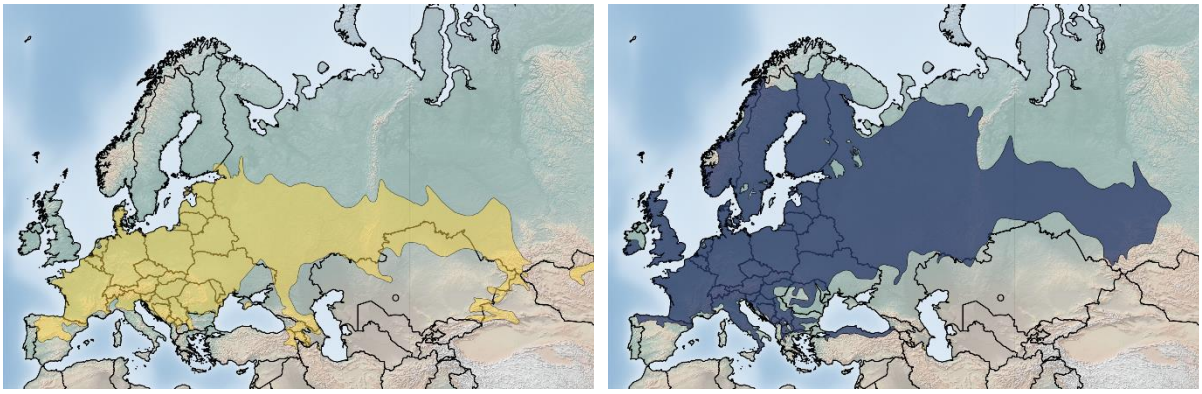


Figure 1.2: Geographic distribution of *M. arvalis* (yellow) and *M. glareolus* (blue) (Shenbrot & Krasnov 2005)

1.6.2 The bank vole (*Myodes glareolus*)

The bank vole *Myodes* (formerly known as *Clethrionomys*, which is an invalid junior name (Carleton et al. 2014)) *glareolus* (Schreber, 1780) (Figure 1.3) has a wide European distribution, ranging from the northern parts of Scandinavia to the southern European regions of the Mediterranean Peninsulas, and from the west coast of France to deep into Russia (Figure 1.2). Within this wide distribution it experiences a range of different climatic and environmental conditions. *M. glareolus'* recent evolutionary history was strongly shaped by the LGM and there is evidence that some populations survived in refugia north of the traditional southern refugia (Kotlík et al. 2006; Bhagwat & Willis 2008; Wójcik et al. 2010). Previous research based on mitochondrial DNA has described seven distinct lineages: The Carpathian, Eastern, Western, Balkan, Spanish, Italian, and Calabrian lineages (Deffontaine et al. 2005; Kotlík et al. 2006; Wójcik et al. 2010; Colangelo et al. 2012; Filipi et al. 2015), which most probably have survived in different glacial refugia and recolonized Europe from there. An eighth "Ural" lineage exists in northern Scandinavia, which was formed by introgression of mitochondrial DNA (mtDNA) from a closely related species, the northern red-backed vole (*Myodes rutilus*) into *M. glareolus* populations (Tegelström 1987; Abramson et al. 2009). The distribution of different lineages might be related to climatic and environmental conditions, as

Figure 1.3: *Myodes glareolus*

suggested by the discovery of possible adaptive mtDNA evolution in *M. glareolus* in the most northern lineages (Filipi et al. 2015), displacement of lineages by other lineages with a presumed selective advantage (Kotlík et al. 2014; Kotlík et al. 2018) and a correlation between lineage distribution and environmental parameters (Tarnowska et al. 2016; Ledevin et al. 2018). There is some evidence for adaptive variation related to climate across its distribution. *M. glareolus* seem to be smaller at higher latitudes, which is thought to increase metabolic efficiency (Ledevin et al. 2010). In Scandinavia, body mass and mass-corrected basal metabolic rate (BMR) decrease with latitude independently of *M. rutilus* mtDNA introgression, indicating adaptive variation along the climate gradient in this region (Boratyński et al. 2011). Genomic resources for *M. glareolus* include several transcriptomes (Konczal et al. 2014; Konczal et al. 2015; Migalska et al. 2017) and a poorly annotated reference genome (GCA_001305785.1). These resources have been used to study the genetic basis of *M. glareolus* physiology in laboratory settings (Konczal et al. 2015; Konczal et al. 2016). In addition to this, studies have been performed to examine the genetic basis of virus tolerance (Rohfritsch et al. 2018) and geographic expansion (White et al. 2013).

1.7 Thesis outline

Despite increasing efforts, we are just beginning to understand the genetic basis of changes that underlie non-model species' abilities to adapt to local environmental conditions. Therefore, the central aim of this work was to shed light on the genetic basis of local adaptation in wild populations of two vole species. Voles are an ideal system to study local adaptation. They have short generation times and large effective population sizes, resulting in high evolutionary potential relative to their capacity to disperse. This induces a high capacity to genetically adapt to their local environmental conditions. Accordingly, I studied different populations along gradients with strong environmental heterogeneity, to try to provide insight into the genetic basis of adaptation to high-altitude in *M. arvalis* populations (Chapter 3) and climate-related adaptive variation in *M. glareolus* populations across Europe (Chapter 4). Additionally, I set out to add genomic resources to facilitate future research in *M. arvalis* (Chapter 2).

Chapter 2

This chapter describes the generation of the first complete mitochondrial genome of the common vole (*M. arvalis*). It was used in phylogenetic analyses together with other species of the Arvicolinae subfamily. Using this mitogenomic dataset, I could confidently place *M. arvalis*

as sister species to *M. levis*, in concordance with other studies (Jaarola et al. 2004; Fink et al. 2010). However, in contrast to studies using only short mitochondrial or nuclear fragments (Galewski et al. 2006), I found the genera *Neodon* and *Lasiopodomys* to fall within the variability of the genus *Microtus*. Thus, our results suggest that these genera should be subsumed within genus *Microtus*

Chapter 3

This chapter sheds light on the genetic underpinnings of high-altitude adaptation in *M. arvalis*. Earlier efforts have been made to identify candidate markers under positive selection related to high altitude in *M. arvalis* (Fischer et al. 2011; Fischer et al. 2014), but the anonymous genetic markers used in these studies hindered obtaining information about the genetic regions or putative genes of interest involved. I therefore aimed to identify loci that show signals of selection associated with differences in altitude. To this end, ddRAD sequencing was used to obtain genetic information from individuals from three low-altitude and three high-altitude study sites in the Swiss Alps. By using genomic resources from a closely related species (*Microtus ochrogaster*), a large number of SNPs shared among individuals was identified. These were used to estimate genetic diversity within populations and examine population structure. I then identified SNPs displaying signals of selection while controlling for the confounding effects of population structure. F_{ST}-based approaches were used to identify SNPs showing signals of local adaptation, and GEA-based approaches were applied to identify SNPs that are associated with differences in altitude. Subsequently, I used the positions of these SNPs on the *M. ochrogaster* genome and identified homologous positions on the *Mus musculus* genome to obtain functional annotation of genes with a signal of selection. Interestingly, a number of genes with functions related to oxygen-transport into tissue and genes related to aerobic metabolic pathways were identified. These findings suggest that high-altitude populations of *M. arvalis* have adapted by enhancing oxygen delivery and thermogenic capacity as a response to the specific conditions at high altitude.

Chapter 4

The final case study in this thesis describes adaptive variation related to environmental gradients in *M. glareolus*. Similar to chapter 3, ddRAD sequencing was used to obtain genomic information from individuals and identify genes to gain insight into the genetic basis of local

adaptation. To this end, individuals from 12 populations distributed across the European range of *M. glareolus* were sampled. I identified a total of 21,892 SNPs and used GEA-based approaches that control for confounding effects of population structure, as well as redundancy analyses (RDA), to identify SNPs with allele frequencies correlating with variation in climate. Additionally, RDA was used to identify the amount of genomic variation related to population structure and climate. The results suggest that both population structure and climate explain similar amounts of genetic variation, but that their effects are mostly shared. This makes it difficult to separate the effects of population structure and climate. However, both GEA-based methods and RDA identified SNPs that show signals of selection related to variation in climate. Among these SNPs, I found a large number located in genes with a function associated with energy homeostasis and immune system functioning, suggesting that variation in temperature and pathogen presence are important selective pressures that drive local adaptation in *M. glareolus*.

Chapter 5

Finally, Chapter 5 provides a general discussion of the results presented in this thesis. It discusses the methods applied in this thesis and illustrates how these can be used to study adaptive variation in non-model organisms with little genomic resources. Moreover, it highlights the general implications of this thesis for evolutionary biology and the study of the genetic basis of local adaptation in particular.

Chapter 2: Article I

The complete mitochondrial genome of the common vole, *Microtus arvalis* (Rodentia: Arvicolinae)*

Remco Folkertsma, Michael V. Westbury, Jana A. Eccard and Michael Hofreiter

This manuscript was published in "Mitochondrial DNA Part B" in 2018. Online available under http://doi.org/10.1080/23802359.2018.1457994

2.1 Abstract

The common vole, *Microtus arvalis* belongs to the genus *Microtus* in the subfamily Arvicolinae. In this study the complete mitochondrial genome of *Microtus arvalis* was recovered using shotgun sequencing and an iterative mapping approach using three related species. Phylogenetic analyses using the sequence of 21 arvicoline species places the common vole as a sister species to the East European vole (*Microtus levis*), but as opposed to previous results we find no support for the recognition of the genus *Neodon* within the subfamily Arvicolinae, as this is, as well as the genus *Lasiopodomys*, found within the *Microtus* genus.

2.2 Main text

The genus *Microtus* is one of the most diverse genera in the subfamily Arvicolinae. Consisting of about 70 different species it is one of the fastest radiating mammalian genera (Nowak 1999). The common vole (*Microtus arvalis*) is a small rodent widely distributed throughout Eurasia, ranging from the Atlantic coast in France to Central Russia. They experience a range of different climatic conditions from sea level to high altitude in the Alps (Fischer et al. 2011) and occupy a variety of different habitats such as farmland and grassland, making it a popular study species in ecological and evolutionary research. The species shows high levels of between population genetic differentiation (Heckel et al. 2005) and strong genetic clustering among populations on small scales (Schweizer et al. 2007). Although previous studies have used partial mitochondrial DNA sequences to show the presence of five main evolutionary lineages and to resolve their phylogenetic positioning within the genus *Microtus* (Fink et al. 2004), the complete mitochondrial genome of *M. arvalis* has not yet been published.

The male *M. arvalis* used in this study was sampled in 2015 outside the town of Lochow, Germany (52.690640°N, 12.455182°E) under permits from Landesumweltamt Brandenburg (RW-7.1 24.01.01.10). A voucher specimen was deposited at the University of Potsdam, Potsdam, Germany. DNA was extracted using the Qiagen DNeasy kit, built into an Illumina sequencing library and sequenced using an Illumina NextSeq 500. We assembled the full mitochondrial genome using an iterative mapping approach (Hahn et al. 2013) with three independent runs utilizing available Arvicolinae as reference bait sequences. Resultant sequences were aligned using Mafft v7.271 (Katoh & Standley 2013) and a final consensus sequence was built using Genious v9.0.5 (Kearse et al. 2012). We obtained a circular sequence, 16,286 bp in length (GenBank Accession No. MG948434) which was annotated using MITOS (Bernt et al. 2013). Finally, a phylogenetic analysis was performed on an alignment of our consensus sequence, all available complete mitochondrial arvicoline sequences and *Cricetulus griseus*. We produced a maximum-likelihood phylogenetic tree of all 13 protein-coding genes and tRNA genes, with an appropriate partitioning scheme and GTR + G as the substitution model as determined by PartitionFinder (Lanfear et al. 2012), with 1000 bootstrap replicates using RAxML-HPC2 on XSEDE v8.2.10 (Stamatakis 2014) on the Cipres server (Miller et al. 2010) (Figure 2.1).

Phylogenetic analysis shows a well-supported sister-clade relationship between Arvicolini and Myodini. Furthermore, in concordance with other studies using mitochondrial or nuclear loci (Jaarola et al. 2004; Fink et al. 2010), we also find *M. arvalis* to be a sister species of *M. levis*. However, in contrast to Galewski et al. (2006) who used the mitochondrial *Cytb* and nuclear *Ghr* genes, we found no support for the recognition of the genus *Neodon*, as this is found within the genus *Microtus*, as is the genus *Lasiopodomys*. Thus, mitogenomic analysis suggests that these genera should be subsumed within the *Microtus* genus. We hope publication of the mitochondrial genome of *M. arvalis* will help to understand the phylogenetic relationship within the Arvicolinae and the genus *Microtus*

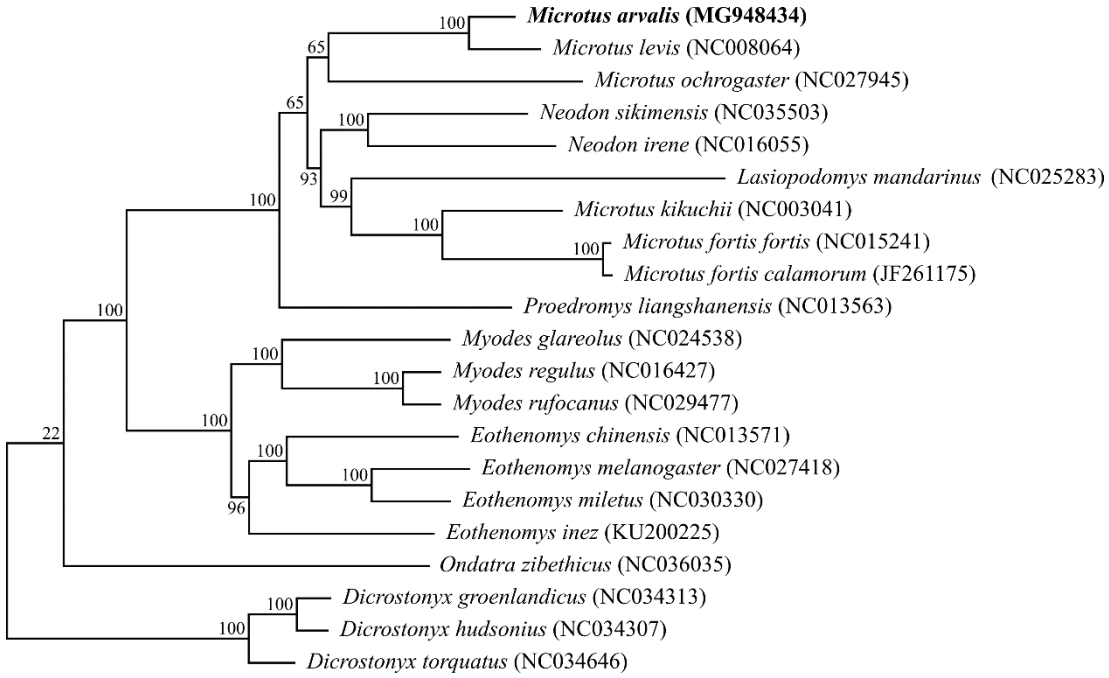


Figure 2.1: Maximum-likelihood tree of the subfamily Arvicolinae based on the sequences of 13 protein-coding genes and the tRNA genes of the mitochondrial genome using *Cricetulus griseus* as an outgroup (not shown here for graphical reasons). Bootstrap support values are shown at the branch nodes.

Chapter 3: Article II

Population genomic evidence for high-altitude adaptation in common voles*

Remco Folkertsma, Jana A. Eccard, Gerald Heckel, Johanna L.A. Paijmans and Michael Hofreiter

*This manuscript has been submitted to “Molecular Ecology” in December 2019.

3.1 Abstract

Local environmental selection pressures can cause adaptive differentiation at the genomic level. Altitudinal gradients have been used to discover adaptive variation in a number of species, as their steep gradients allow examining signatures of selection in the face of gene flow. Studies using small mammals have offered valuable insights into the genetic basis of adaptation to high-altitude environments. They experience high levels of thermogenic and aerobic stress as a result of cold and hypoxic condition with increasing altitude and, because of their small size, are less likely to migrate between habitats at different altitudes. Here, we used ddRAD-sequencing and landscape genomic approaches to examine population structure and evidence for adaptation to high-altitude environments in six populations of *Microtus arvalis* from low and high-altitude study sites in the Swiss Alps. We discovered high levels of genetic differentiation between study sites and strong patterns of population structure with a signal of isolation by distance. We used F_{ST}-based and genotype-environment association (GEA)-based methods that control for the confounding effects of population structure, to identify loci showing signals of selection associated with differences in altitude. Across all different methods, we identified 127 candidate loci potentially related to high-altitude adaptation, 46 of which were within 37 genes. Functional annotation of these genes suggests that selection in high-altitude populations of *M. arvalis* is acting to increase oxygen delivery to tissue and on aerobic metabolic pathways in responses to cold hypoxic conditions. Several of the candidate genes have been linked with adaptation to high altitude before, suggesting a rather specific effect of selection pressures at high altitude on mammalian genomes.

Keywords: Common vole, High altitude, Genome scan methods, ddRAD, adaptation, *Microtus arvalis*

3.2 Introduction

A key goal of research in evolutionary ecology is to understand the genetic basis of adaptive differentiation of populations in response to environmental variation (Storz 2005; Hereford 2009). Natural selection shifts allele frequencies and may thus lead to genetic differences between locally adapted populations (Richardson et al. 2014). Local adaptation is assumed when populations in their native environment have higher fitness than populations of the same species from any other environment (Kawecki & Ebert 2004; Savolainen et al. 2013). Together with spatially varying selection, genetic drift and gene flow are the main factors influencing genetic population structure across the landscape (Hedrick 2011). Both high levels of gene flow and strong genetic drift can diminish local adaptation. Gene flow can hinder local adaptation by introducing alleles from differentially adapted neighboring populations, thereby swamping locally favorable genes (Lenormand 2002). However, it can also facilitate local adaptation by introducing genetic variation to act upon and alleviate the effects of genetic drift (Garant et al. 2007; Savolainen et al. 2007). If the strength of selection is high enough and the amount of gene flow is not too high, divergence between populations at adaptive loci can be maintained while other parts of the genome will be homogenized by the effect of gene flow (Feder et al. 2012). Genetic drift can also cause loss of some and fixation of other alleles, leading to structure in genetic variation possibly looking similar to that caused by local adaptation (Lande 1976; Yeaman & Otto 2011).

Environments with spatially varying selection pressures offer the opportunity to study the components that form population genetic differentiation and local adaptation (Haldane 1948; Storz 2002). The field of landscape genomics provides a framework for identifying adaptive genetic variation in a spatial context and with the recent decrease in sequencing costs and the accompanying increased availability of genome-scale genetic data, the opportunity arose to also study non-model organisms (Schoville et al. 2012; Rellstab et al. 2015; Haasl & Payseur 2016). Genome scans trying to identify adaptive genetic variation can be broadly defined in two categories. Differential outlier methods (F_{ST}-based methods) seek for signals of selection by identifying alleles that show genetic differentiation that is unusual high (evidence for directional selection) or low (evidence for balancing selection) among populations (Beaumont & Balding 2004; Foll & Gaggiotti 2008; Luu et al. 2017). In contrast, genotype-environment association (GEA)-based methods identify loci with allele frequencies that are strongly associated with environmental variables of interest, and those loci are interpreted as potentially under positive

selection (Frichot et al. 2013; Günther & Coop 2013; Gautier 2015). With these tools, several studies using non-model organisms have used environmental gradients to infer local adaptation to climate (Branco et al. 2017; Harrisson et al. 2017), urbanization (Harris & Munshi-South 2017), or salinity (Guo et al. 2015). Altitudinal gradients are particular suitable to study selection pressures imposed by environmental variables, as they are often steep and occur at relatively small geographic scales. Therefore, local conditions can vary dramatically over short spatial distances (Körner 2007). This potentially offers the benefit of studying adaptive differentiation in the face of gene flow and without having the confounding effect of populations’ individual demographic histories (Keller et al. 2013). The most obvious environmental changes along altitudinal gradients are decreasing atmospheric pressure and oxygen levels (hypoxia), decreasing temperature, and increasing solar radiation intensity with increasing altitude (Körner 2007). Using genome-wide data and population genetic methods, genetic evidence for physiological adaptation to high altitude have been found in a range of mammals such as humans, pigs, sheep, goats and dogs (e.g. Ai et al., 2014; Gou et al., 2014; Simonson et al., 2010; Song et al., 2016; Wei et al., 2016).

It has been suggested that small mammals experience particularly strong selection pressure caused by the physiological challenges in high-altitude environments. With their small body size and proportional large body surface they have to sustain high rates of aerobic metabolism to support thermogenesis and locomotion, which is particularly challenging at high altitude under cold conditions with only limited levels of oxygen available (Tucker 1970; Hayes 1989). For example, high-altitude populations of deer mice (*Peromyscus maniculatus*) evolved physiological differences in aerobic exercise performance and thermogenic capacities in hypoxic conditions compared to their lowland counterparts (Cheviron et al. 2012; Lui et al. 2015). Genetic changes associated with these differences are related to metabolic pathways involved in oxygen transport and oxidative capacity of mitochondria, suggesting an adaptive response to cope with lower air temperatures and lower oxygen levels (Lui et al. 2015; Scott et al. 2015; Lau et al. 2017; Mahalingam et al. 2017). Similar, the mitochondrial haplotype of white-toothed shrew (*Crocidura russula*) is associated with non-shivering thermogenesis along an altitudinal gradient, also suggesting selection is acting on thermogenic capacity (Ehinger et al. 2002; Fontanillas et al. 2005). Finally, two independent studies of pika species living at high altitude (*Ochotona princeps* and *O. roylei*), identified regulatory changes in genes involved in metabolic function and oxygen transport that play a role in coping with low oxygen levels

(Waterhouse et al. 2018; Solari et al. 2018). These results suggest that among small mammals living at high altitude, common adaptive responses exist that help them survive in these cold and hypoxic conditions.

The common vole, *Microtus arvalis* (Pallas, 1778), is one of the most abundant mammal species in Europe. Its distribution ranges from the Atlantic coast of France to the steppes of Central Russia, and from the Baltic sea coast in the North to the Mediterranean coast in Spain. In altitude, it ranges from sea level to above 2,000 m a.s.l (Fischer et al. 2011). It lives in dry habitats such as fields and meadows, where it builds subterranean burrows, feeding on plants, seeds and occasionally insects. Its life-history is characterized by a short lifespan and early maturation. Females show high levels of fecundity and can produce up to 5 litters per reproductive season (Boyce & Boyce 1988). Voles have a low and sex dependent effective dispersal rate. After maturation, females usually stay in close proximity to their burrow, while males show a high tendency to disperse among colonies (Schweizer et al. 2007; Borkowska et al. 2010; Hahne et al. 2011). This leads to strong genetic clustering at large (Heckel et al. 2005; Fischer et al. 2014) and small (Schweizer et al. 2007; Borkowska et al. 2010) geographic scale. As opposed to many other species during the last glacial maximum (LGM), *M. arvalis* had surviving populations not only in the south in refugia located on the Balkan and the Iberian Peninsula (Hewitt 1999), but also survived in locations further north in Central and Western Europe (Fink et al. 2004; Hamilton et al. 2005; Pedreschi et al. 2019). However, colonization of the central regions of the Alps started only as recently as 10,000 years before present, as this area was covered with ice during the LGM (Hamilton et al. 2005; Braaker & Heckel 2009). This suggests that *M. arvalis* experienced higher selection pressures in its more recently colonized high-altitude environments compared to longer-term inhabited low-altitude environment.

Evidence for local adaptation associated with altitude comes from two previous studies. By comparing populations living at different altitudes, Fischer et al. (2011) identified signatures of local adaptation associated with high altitude in four Alpine populations using AFLP markers (1.25% of markers). A second study, on a larger European scale, uncovered that about 6.7% of the studied AFLP markers were under positive selection, and geographic patterns suggest that many of these are associated with Alpine regions favoring adaptation as a result of their extreme environmental conditions (Fischer et al. 2014). Unfortunately, the use of anonymous markers

in these studies prevented them from obtaining information about the genetic regions or putative genes involved in local adaptation of *M. arvalis*.

Here, we employed a double digest restriction-site associated DNA (ddRAD) sequencing approach to generate tens of thousands of SNPs in *M. arvalis* across six study sites sampled at high and low altitude in the Swiss Alps. Our primary aim was to identify candidate genes involved in evolutionary adaptation to high altitude. To this end, we used a set of outlier methods to identify loci with signatures of positive selection and specifically searched for signals of divergent selection between high and low altitude populations, while accounting for population structure to avoid detection of false positives.

3.3 Materials and methods

3.3.1 Sampling, library preparation and sequencing

We collected a total of 143 *M. arvalis* from three low-altitude (~ 600 m a.s.l.) and three high-altitude sites (~2.000 m a.s.l.) in the Swiss Alps. Low-altitude sites are referred to as L1W; L2W; and L3E, high-altitude sites are referred to as H1; H2; and H3 (Figure 3.1, Table 3.1). Study sites are in close proximity to each other and are located within the distribution of the Central evolutionary lineage of *M. arvalis*, with at least 70 km distance to the nearest contact zone or populations carrying mitochondrial DNA from other lineages (Braaker & Heckel 2009; Beysard & Heckel 2014).

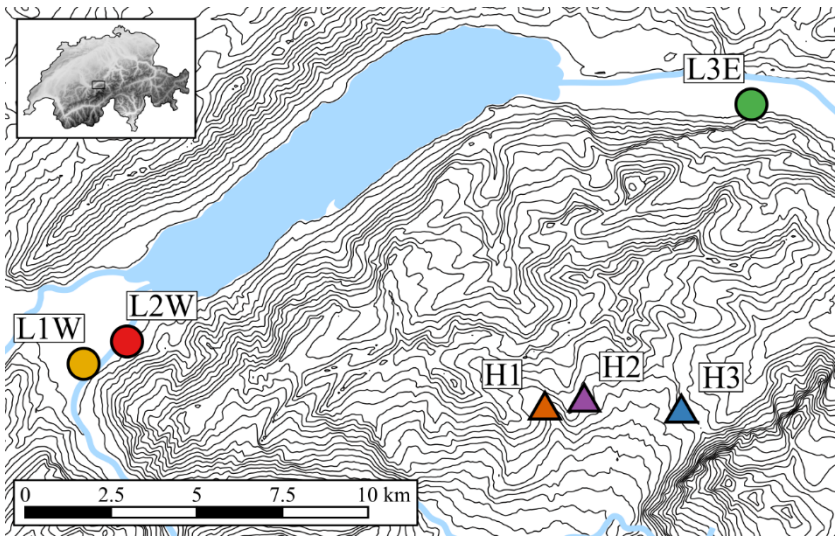


Figure 3.1: Map of study sites of *M. arvalis* populations in the Swiss Alps. Circles indicate low-altitude (~ 600 m a.s.l.) and triangles high-altitude sites (~2,000 m a.s.l.). Codes for study sites are in Table 3.1. Contour lines represent 100 m change in altitude, Lake Brienz and streams are indicated in blue.

Voles were trapped using snap traps and tissue samples were preserved in absolute ethanol and stored at -20°C. Samples were then analyzed using a double digest restriction associated DNA (ddRAD) sequencing protocol (Peterson et al. 2012). Briefly, total DNA was extracted from one paw of each individual (approximately 15 mg of tissues) using the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer’s recommendations and quantified using a Qubit 2.0 Fluorometer (Life Technologies Inc., ON, Canada). As DNA degradation can cause a loss of data when using ddRAD (Graham et al. 2015), DNA extracts were visually checked for degradation using 1% agarose gels and only non-degraded extracts were selected for further processing. For each individual, we independently digested 1 µg of DNA in 50 µl reactions for 12 hours at 37ºC with 20 units each of EcoRI-HF and MspI-HF enzymes (New England Biolabs, NEB). Digests were cleaned using Sera-Mag Speedbeads (Fisher Scientific, Pittsburgh, PA, USA) and eluted in 40 µl of H_2O. DNA concentrations were quantified using the QuantiFluor dsDNA systems (Promega) and each sample was standardized to 125 ng of digest before carrying out a ligation reaction. During this reaction, samples were ligated to P1 adapters that contain a unique 5-nucleotide barcode (designed to ligate to the sticky end left by EcoRI) and a P2 forked-adapter (designed to ligate to the MspI-end). Ligation reactions were performed at 25°C for 1 hr, after which the enzyme was heat-inactivated at 65°C for 10 min, followed by a decrease of 2°C every 90 s until the reaction reached room temperature. PCR products for each individual were visualized on an agarose gel to ensure ligation and amplification. Ligated samples were then grouped into pools of 36 individuals and cleaned using Speedbeads. Each pool was size selected for fragments 300-400 bp in length using a pippin prep system (Sage Science, Beverly, MA). This range was expected to provide ~ 65.000 ddRAD loci based on in silico digestion of the prairie vole (*M. ochrogaster*) genome (accession: GCF_000317375.1) using SimRAD (Lepais & Weir 2014). We confirmed the sizes on an Agilent Tapestation 2200. Next, we performed a qPCR to determine the optimal number of cycles for each pool, based on the start of the saturation phase on amplification plots (range: 12-14 cycles) (Gansauge & Meyer 2013). Pools were then amplified by PCR in four parallel reactions of 40 µl using primers designed to only amplify fragments with P1 and P2 adapters according to the protocol by Peterson et al. (2012). The resulting libraries were sequenced in two separate runs on an Illumina NextSeq 500 using mid-output kits. We first sequenced libraries using 75 bp paired end (PE) sequencing after which we performed another sequencing run using 150 bp PE sequencing.

3.3.2 Mapping

Following sequencing, raw sequences were treated with cutadapt v1.4 (Martin, 2011) to remove Illumina-specific adapters and trim low-quality bases from reads at their 3' end with a phred score below 20. Reads were demultiplexed based on the sequence of their internal barcode not allowing for any mismatch using iPyrad v0.7.13 (Eaton 2014). Only reads with a minimum length of 35 bp were retained. Remaining reads were then mapped onto the *M. ochrogaster* reference genome using bwa-mem v0.7.8 (Li, 2013) with default parameters. We decided to map against a closely related genome rather than using a *de novo* approach, as this has been shown to provide better allele frequency estimates (Shafer et al. 2017). We then converted mapping results in SAM files into BAM files using SAMTOOLS v1.6 (Li et al., 2009), while filtering for a minimum mapping quality of 30 and only kept uniquely mapping reads in proper pairs. Finally, we performed local realignment around indels using GATK's v3.8 (McKenna et al. 2010) RealignerTargetCreator and IndelRealigner (DePristo et al. 2011) to avoid incorrect SNPs associated with misaligned reads.

3.3.3 Genotype calling and genotype likelihoods

Genotype calling based on low to moderate coverage data is associated with uncertainties that can lead to biased estimates of allele frequencies and population genetic parameters. Therefore, genotype likelihoods were estimated using ANGSD v0.914 (Nielsen et al. 2012; Korneliussen et al. 2014). This software, specifically designed for such data, estimates genotype likelihoods while taking the uncertainty caused by errors in base calling, mapping and alignment into account. Moreover, ANGSD has the possibility to determine the major and minor allele, respectively, based on the genotype likelihoods, which is an advantage when aligning reads using a divergent reference genome (Skotte et al. 2012). This provides the best estimate of variability when a species-specific reference genome is not available (Nevado et al. 2014). We used the estimated genotype likelihood and then called genotypes using the following filters: A minimum base quality of 20; a maximum of 2 alleles; a minor allele frequency of 0.05 across all samples (minor/major alleles were inferred from the genotype likelihoods); a maximum p-value threshold of 10^{-6} for calling a SNP; a minimum depth of 5 and a maximum depth of 50 per sample and a genotype posterior probability cut-off value of 0.95. Genotypes were only retained if they could be genotyped in at least 70% of all individuals and in at least 12 individuals from each study site (Huang & Knowles 2016). Finally, genotype calls were output in PLINK (Purcell et al. 2007) and VCF (Danecek et al. 2011) format and used in subsequent

genome scan methods. To avoid bias and errors in downstream analyses that can be caused by uncertainty in called genotypes (Johnson & Slatkin 2008; Han et al. 2014), we estimated population genetics parameters and examined population structure based on the likelihood of site frequency spectra (SFS), genotype likelihoods and genotype probabilities. For this, we created a second dataset containing only genotype likelihoods. We used the same filters as described above except that we only included SNPs with a minimum read-depth of 2, we did not allow for missing individuals, and included only one SNP per RAD-tag.

3.3.4 Population diversity measurements

We estimated population genetic diversity for individuals from each study site based on their SFS, which was polarized using the reference as ancestral state (Korneliussen et al. 2013). Nucleotide diversity was first calculated as the average number of pairwise differences (π) (Nei & Li 1979) and second as the proportion of segregating sites (θ_W) (Watterson 1975). From these estimates, we subsequently calculated Tajima's D (Tajima 1989). In a similar way, we estimated the SFS for each individual separately to calculate individual level heterozygosity. Per-individual inbreeding coefficients (*F*) were estimated using ngsF (Vieira et al. 2013). Further, to infer relationships between individuals, we estimated the relatedness coefficient (*r*) between pairs of individuals using NgsRelate (Korneliussen & Moltke 2015). NgsRelate calculates *r* by obtaining the maximum-likelihood coefficients for K_0, K_1 and K_2, the probability that two individuals share 0, 1 or 2 alleles by descent.

3.3.5 Population structure

To infer population structure and admixture among populations we used several methods. First, we ran NGSadmix (Skotte et al. 2013) with the number of clusters *K* ranging from two to eight to account for possible within-site structure. Each value of *K* was repeated 20 times or until three replicates were within one likelihood unit of the highest value. All values of *K* converged using these criteria; we therefore only report the runs with the highest likelihood here. As an alternative approach, population structure and relationships between individuals were visualized using a principal component analysis (PCA). For this, we approximated the covariance matrix among individuals using ngsCovar from the ngsTools suite (Fumagalli et al. 2014). We then used the 'eigen' function in R v3.4.4 (R Core Team 2018) to calculate principal components and produced PCA plots. Furthermore, we estimated the weighted (ratio of sums) pairwise F_{ST} using the Hudson estimator (Hudson et al. 1992; Bhatia et al. 2013), based on the

two-dimensional SFS for each population pair. We tested for patterns of isolation by distance (IBD) by testing the correlation between pairwise linearized F_{ST} values [$F_{ST}/(1-F_{ST})$] and the log-transformed geographic distance (Rousset 1997), using a Mantel test with 1,000 permutations in the R package 'Vegan' (Dixon 2003).

3.3.6 Identifying loci under selection

We used different genome-scan methods with the called genotypes dataset in order to identify SNPs likely to be differentiated as a result of selection. A major problem for such methods is to clearly distinguish between genetic signals left by neutral processes and selection, respectively. Therefore, it is recommended to combine multiple methods as this reduces error rates and false positives (de Villemereuil et al. 2014; François et al. 2016).

First, to detect loci which are highly differentiated between populations, we implemented two F_{ST}-based methods. Bayescan v2.1 (Beaumont & Balding 2004; Foll & Gaggiotti 2008; Fischer et al. 2011) identifies SNPs that show extreme F_{ST}-values between populations compared to what would be expected under neutral processes. We ran three replicates using the software’s default values, using a prior odds value of 100 and a q-value cutoff of 0.1 (corresponding to a false discovery rate (FDR) of 10%). SNPs were considered outliers if they were identified in each replicate. BayeScan requires individuals to be grouped into populations *a priori*. In our analysis, individuals were grouped into 6 populations according to study site. Next, we used Pcadapt v3.1 (Luu et al. 2017), a principal components-based method that does not need *a priori* grouping of individuals into populations. Instead, Pcadapt requires users to estimate the correct number (K) of principal components that need to be retained in order to correct for neutral population structure. We decided upon K=5, after first running Pcadapt with K=1-20 and based on visual inspection of the resulting scree plot using Cattel's rule (Cattell 1966), as recommended by the authors. SNPs with a q-value < 0.1 were considered outliers. We also used two complementary GEA-based methods to identify outlier SNPs that are associated with differences in altitude (coded as 1 and -1 for high and low-altitude populations, respectively).

First, we used BayPass (Gautier 2015), which accounts for demographic history of populations by computing a population covariance matrix based on the population's allele frequencies and then tests for non-random association between covariables and population allele frequencies using a Bayesian framework. The BayPass model is an extended and improved version of the

Bayenv model proposed by Coop, Witonsky, Di Rienzo, & Pritchard (2010) and Günther & Coop (2013), which improves estimation of the population covariance matrix and includes a complete reprogramming of the Markov Chain Monte Carlo (MCMC) algorithm. We first estimated the population covariance matrix using the core model with 20 pilot runs of 1,000 iterations, followed by a burn-in of 50,000 iterations and an MCMC procedure of 50,000 iterations, sampling every 25 iterations. Then we ran BayPass in the auxiliary covariate model with the same settings, while using the previously estimated covariance matrix to identify loci associated with altitude. We considered individuals to come from three populations as this best represents the population structure as identified by principal component analysis and admixture analysis (see Results section). We performed five independent runs under the auxiliary model to check for consistency of the estimates and then calculated the median Bayes Factor (BF) for each SNP. As a decision rule for association of a SNP with the low/high altitude status of populations, SNPs with a BF > 15 decibans, suggesting very strong evidence for selection according to the Jeffreys scale (Jeffreys 1961), were considered as outlier SNPs under selection.

Second, we used latent factor mixed models (LFMM) to test for genotype-environment associations using the LEA R-package (Frichot & François 2015). LFMM uses a univariate mixed-model approach that estimates the correlation between allele frequencies and environmental values, while taking into account the neutral genetic background and levels of ancestral population structure by using so-called latent factors, which represent the genetic structure of the data (Frichot et al. 2013). To define the appropriate number of latent factors, we used the snmf function to determine the number of hidden genetic clusters (K). We identified the best value of K by comparing the cross-entropy score in the snmf analyses for each K between 1 and 15 with 20 repetitions each and chose the proper K-value where there was no further strong decline in cross-entropy score. We then performed 10 independent runs of LFMM to test for association of allele frequencies with altitude, using 50,000 burn-in cycles followed by 100,000 iterations, with K=3. From these we calculated the median z-scores and then readjusted the p-values using the genomic inflation λ, calculated as median(z-scores^2)/0.456 (Devlin & Roeder 1999). Finally, we obtained a list of outlier SNPs after controlling for multiple testing using the Benjamini-Hochberg procedure (Benjamini & Hochberg 1995) with an expected FDR equal to 10%. Outlier SNPs detected by at least two methods were considered candidate SNPs for further analysis (de Villemereuil et al. 2014).

3.3.7 SNP annotation and gene ontology

We obtained annotations for each SNP using the annotation release 101 provided with the *M. ochrogaster* reference genome. We then used biomart (Smedley et al. 2015) and the Ensembl genome database to find homologous *Mus musculus* genes for each *M. ochrogaster* gene and used these to retrieve associated gene ontology (GO)-terms using the UniProt knowledgebase (Bateman et al. 2017). To evaluate over-represented terms of biological process ontologies, we used the Bioconductor package topGO (Alexa & Rahnenführer 2010), to compare our list of candidate SNPs with all identified genes with a SNP. We used a Fisher's exact test and the *elim* algorithm to account for correlation in the GO graph topology, and report those GO-terms with a p-value < 0.01 and at least 3 associated genes.

3.4 Results

3.4.1 Sequencing results and filtering

We used a ddRAD-sequencing approach for genome-wide genotyping of 143 individuals of *M. arvalis* from six locations in the Swiss Alps. In total, we obtained 93,862,985 reads during two runs of sequencing. After filtering for low quality reads and assigning individuals to barcodes, this resulted in an average of 654,229 (SD = 227,696) high-quality PE reads per individual ranging from 14.1 to 17.1 million reads per study site. Of these, approximately 83.4 % (SD=2.2%) aligned to the *M. ochrogaster* reference genome. Amongst individuals, mapped reads covered an average of 14,158,248 (SD=524,786) nucleotides of the genome (~0.62%) with an average sequencing depth of 5.60X (SD=1.50) ranging from 2.69X to 10.48X per individual. After filtering, we obtained genotype information for a total of 19,119 SNPs, corresponding to 9,318 RAD-tags, that can be used in downstream genome scan analyses. In addition to this we obtained genotype likelihoods for 2,358 SNPs for use in population genetic analyses. These were calculated for independent SNPs that contained genotypic data for all individuals.

3.4.2 Genetic diversity statistics

First, we estimated genetic diversity for each of the six study sites (Table 3.1). The number of pairwise differences and the number of segregating sites was very similar for each location and did not differ significantly between high and low-altitude sites, although all estimates were slightly lower in high-altitude sites. The genome wide average Tajima’s D was positively skewed for all sites, ranging from 0.56 to 0.78 and did not differ significantly between high and

low-altitude sites. Inbreeding coefficients (*F*) were generally low, but varied among populations. Results from NgsRelate suggest varying degrees of relatedness of individuals within populations. Except for three individuals from *H2*, which are related to an individual from *H1*, we did not find evidence for relatedness between individuals from different study sites (Appendix A: Table S3.1). Individual levels of heterozygosity were generally high and there was no significant difference among populations, except for *H3*, which was lower in all pairwise comparisons with low-altitude sites (Tukey's HSD tests; $p < 0.05$).

Table 3.1: Overview of characteristics and genetic diversity statistics (mean ± standard deviation) of *M. arvalis* at six study sites from low and high altitude in the Swiss Alps. Population ID (ID), sample size (n), altitude in metre above sea level, Watterson's theta (θ_W), Tajima's pi (π), Tajima's D, heterozygosity - the proportion of heterozygous genotypes, inbreeding coefficients (F_{IS}) and the kinship coefficient (*r*) are displayed.

ID	Study site	n	Altitude (m)	Latitude	Longitude	θ_W	π	Tajima's D	Heterozygosity	F_{IS}	*r*
Low altitude											
						0.00140	0.0021	0.68	0.00123	0.004	0.04
L1W	Wilderswil	24	580	46.670 °N	7.873 °E	(0.00163)	(0.0027)	(1.10)	(0.00012)	(0.009)	(0.02)
						0.00138	0.0020	0.78	0.00122	0.015	0.04
L2W	Böningen	24	585	46.676 °N	7.887 °E	(0.00149)	(0.0024)	(1.15)	(0.00013)	(0.041)	(0.01)
						0.00137	0.0020	0.77	0.00125	0.000	0.01
L3E	Meiringen	23	580	46.738 °N	8.124 °E	(0.00155)	(0.0025)	(1.12)	(0.00013)	(0.000)	(0.01)
High altitude											
						0.00138	0.0020	0.71	0.00124	0.016	0.05
H1	Waldspitz	24	2,060	46.660 °N	8.046 °E	(0.00155)	(0.0025)	(1.07)	(0.00021)	(0.052)	(0.02)
						0.00132	0.0019	0.56	0.00113	0.014	0.03
H2	Schreck-Feld	24	2,036	46.662 °N	8.062 °E	(0.00168)	(0.0027)	(1.04)	(0.00014)	(0.026)	(0.02)
						0.00119	0.0018	0.78	0.00109	0.003	0.06
H3	Grosse Scheidegg	24	1,980	46.659 °N	8.098 °E	(0.00136)	(0.0023)	(1.16)	(0.00009)	(0.009)	(0.01)

3.4.3 Population structure and differentiation

Next, we examined population structure using the dataset with genotype likelihoods. All analyses showed a general pattern of neutral genetic structure based on the geographic context of study sites. Analyses using NGSadmix revealed a clear pattern of genetic clusters when considering a *K* value of three, with little admixture between genetic clusters and an assignment probability to a single genetic cluster above 90% for each individual. Here, individuals from *L3E* form a single distinct cluster, another cluster is formed by individuals from *L1W* and *L2W*, and a third cluster consists of all high-altitude individuals from *H1*, *H2* and *H3* (Figure 3.2). Higher values of *K* revealed additional substructure associated with geographic location. With *K*=6, individuals are coarsely grouped into clusters coherent with their geographic sampling location, except for individuals from *H1* and *H2*, which are also the geographical closest to each other.

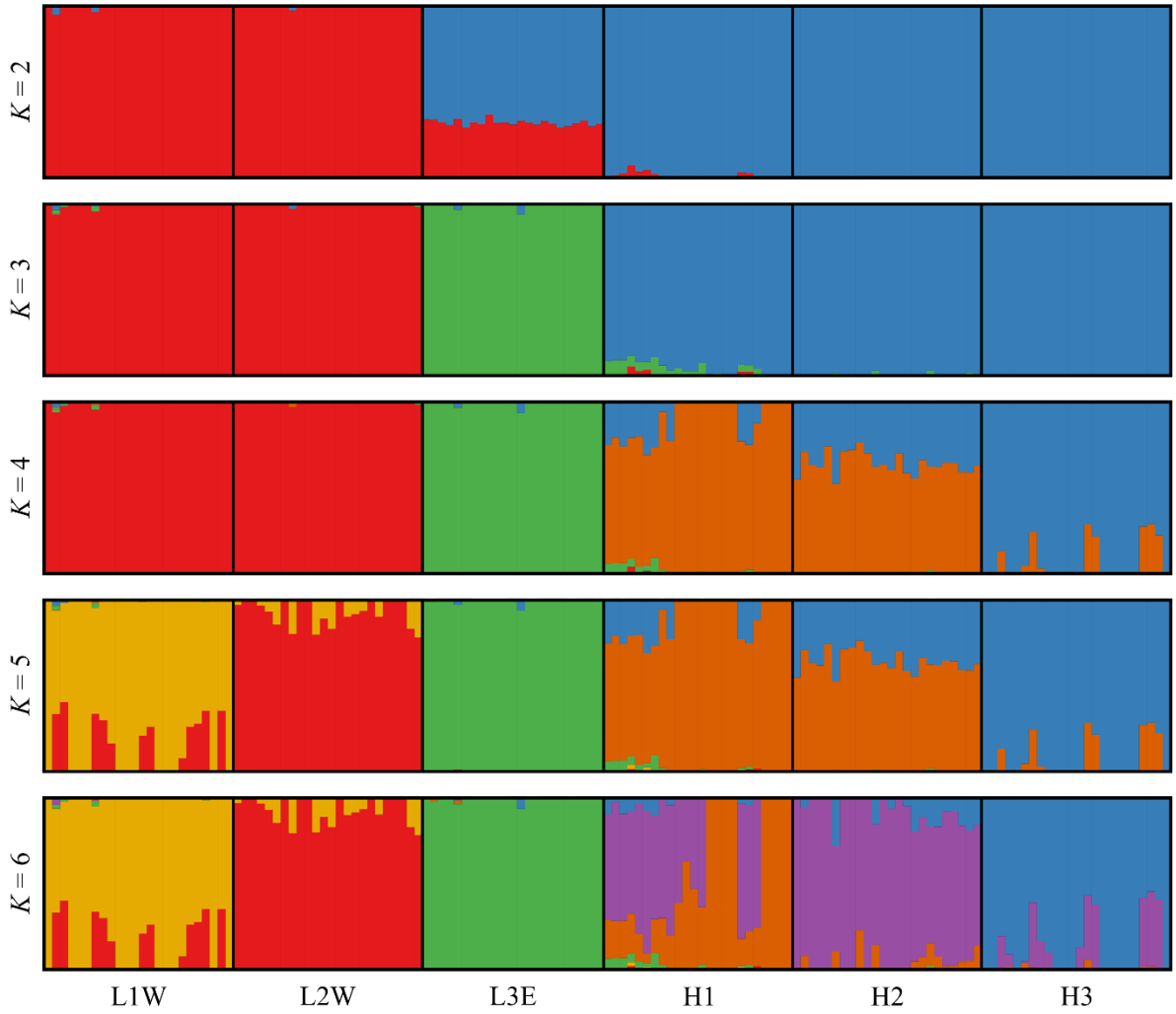


Figure 3.2: Admixture proportions using NgsAdmix based on 2,358 variable sites with various numbers of ancestral populations (*K*= 2-6) of *M. arvalis*. Each individual is represented by a column with colours corresponding to the proportions of their ancestry component. Vertical black bars separate study sites. Codes for study sites are in Table 3.1.

The first two components of the PCA using ngsCovar explain 16.2% of genetic variation and also clearly suggest a similar clustering of individuals based on their geographic location (Figure 3.3). The first axis, which explains 10.9% of genetic variation, shows a cluster of individuals from *L1W* and *L2W* separated from a cluster of individuals from high-altitude study sites of *H1*, *H2* and *H3*. The second axis, which explains 5.3% of genetic variation, shows individuals of *L3E* in a cluster separated from the individuals from the other study sites. Analyses using the called genotypes dataset provide the same general pattern (Appendix A: Figure S3.1 and Figure S3.2). The inferred population structure corresponds well with genomic pairwise F_{ST}-values, which revealed high levels of population differentiation ranging from 0.05 (*H1* vs. *H2*) to 0.32 (*L2W* vs. *H2*). High F_{ST}-values for *L3E* were found in all pairwise

comparisons (all pairwise F_{ST}-values > 0.23) (Appendix A: Table S3.2). Pairwise F_{STS} were positively correlated with differences in geographic distance (Mantel's $r = 0.87$, $p < 0.01$; Appendix A: Figure S3.3a). As this result might be driven by the bimodal nature of the data, we also tested this with only distant sites, however, results remained significant (Mantel's $r = 0.68$, $p = 0.03$; Appendix A: Figure S3.3b). Pairwise F_{STS} did not correlate with differences in altitude (Mantel's $r = 0.82$, $p > 0.1$).

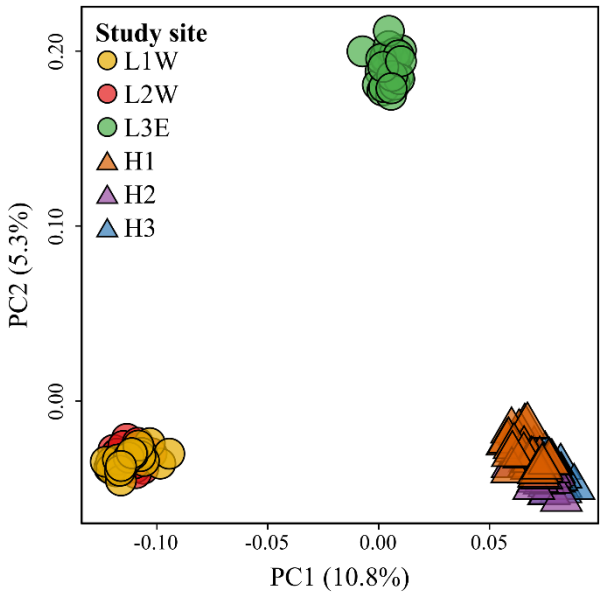


Figure 3.3: Neutral genetic structure of populations as depicted by the first two principal components of a PCA based on pairwise genetic covariance among 143 individuals with 2,358 variable sites of from six *M. arvalis* study sites. Codes for study sites are in Table 3.1.

3.4.4 Candidate loci under selection

In total, all approaches together identified 696 (3.6%) outlier SNPs belonging to 546 unique RAD-tags. First, using F_{ST}-based methods to detect SNPs under divergent selection, we identified a total of 587 SNPs (3.1%) as outliers. BayeScan identified 226 SNPs (1.2%, in line with the findings of Fischer et al. (2011) as outliers using an FDR of 10%, with 80 of these found in genes. None of these outliers were found to be under balancing selection. Analysis with Pcadapt resulted in 430 SNPs (2.2%) detected as outliers, 156 of which were located within a gene. Of all SNPs detected by F_{ST}-based approaches, 69 SNPs (7.3%) were detected by both BayeScan and Pcadapt (Figure 3.4a). Analysis with BayPass revealed that 115 SNPs (0.6%) were significantly associated with altitude using a Bayes factor above 15. Out of these, 44 were located within a gene. LFMM was more conservative, identifying 92 SNPs (0.5%) as associated with altitude using an FDR of 10%, 22 of which were located within a gene. This combined to a total of 181 SNPs (0.9%) detected using GEA-based methods, of which 26 SNPs (14.4%)

were identified by both methods (Figure 3.4b). There was little overlap between SNPs detected using different approaches. First, only 72 SNPs (10,3%) were overlapping between F_{ST}-based methods and GEA-based methods, and only 7 SNPs were detected by all four methods (Appendix A: Figure S3.4). Together these combined to 127 candidate SNPs detected by at least two methods. Of these, 46 were located within genes, representing 37 unique genes. Several of these genes were associated with known function in hypoxia signaling pathways, red blood cell production, angiogenesis and energy metabolism (Table 3.2). Enriched terms for our candidate SNPs tended to be rather general, but two interesting terms related to circadian rhythm and glomerulus development, respectively, were also detected (Appendix A: Table S3.3).

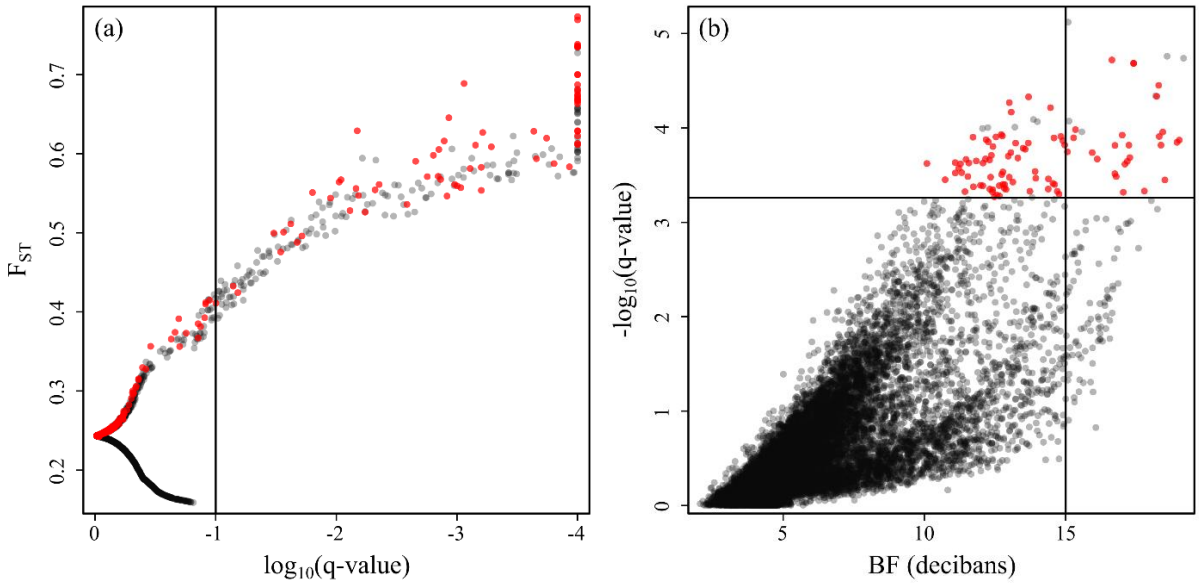


Figure 3.4: Results of outlier detection based on 19,119 SNPs in six *M. arvalis* study sites. (a) BayeScan plot with F_{ST} on the y-axis plotted against the log10 of the q-values. The vertical line indicates the cut-off (FDR=10%) used for detecting outliers with BayeScan, detected SNPs are on the right side of this line. Red circles indicate SNPs detected as outliers by Pcadapt. (b) Results of the genotype-environment association analyses with low-altitude populations coded as -1 and high-altitude populations coded as 1. Depicted is the correlation between Bayes Factor (in decibans) from the BayPass analysis and the -log10 of q-values from the LFMM analysis. The vertical line represents the BF > 15 deciban decision rule for detected outliers with BayPass. SNPs on the right side of this are detected as outliers. The horizontal line represents the value of the Benjamin-Hochberg correction threshold for outliers detected with an FDR of 10% with LFMM. Circles above this line in red are detected as outliers by LFMM.

Table 3.2: Subset of top candidate loci identified with at least two genome scan methods. Locus refers to the location on the *M. ochgrogaster* reference genome (Chromosome_base-position). Method of outlier detection refers to the genome scan method that identified the SNP as an outlier (LFMM=LFMM, BP= BayPass, BS=BayeScan, PCA=Pcadapt).

Locus	Gene ID (ENSEMBLE)	Gene (abbreviation and synonyms)	Method of outlier detection	References	Putative function and relevance
NC_022012.1_70399315	ENSMOCG00000016886	RAR related orphan receptor A (Rora)	LFMM, BS	(Kim et al. 2008; Alkorta-Aranburu et al. 2012)	Involved in hypoxia signaling pathways, induces transcription of HIF-1a , associated with variation in highland Ethiopians hemoglobin levels
NC_022012.1_76422015	ENSMOCG00000022058	Mitogen-activated protein kinase kinase 5 (Map2k5)	BP, BS	(Chen et al. 2007; Xing et al. 2013)	Regulator of angiogenesis, under selection in high-altitude Mongolians
NC_022012.1_93309041	ENSMOCG00000013053	Unc-51 like kinase 4 (Ulk4)	BP, BS	(Levy et al. 2009)	Associated with blood pressure and hypertension in humans
NC_022015.1_4954291	ENSMOCG00000010413	Aryl hydrocarbon receptor nuclear translocator like (Arntl/Bmal1)	BP, BS, PCA	(Ghorbel et al. 2003; Jensen et al. 2012)	Circadian clock gene involved in the HIF-pathway, influencing metabolic adaptation to hypoxia and angiogenesis
NC_022016.1_35085089	ENSMOCG00000018101	GATA binding protein 1 (Gata1)	BS, PCA	(Pevny et al. 1991; Zhang et al. 2012)	Transcriptional regulator of red blood cell production, increased expression induced by HIF-1a under hypoxic conditions
NC_022016.1_52793831	ENSMOCG00000022373	Erythrocyte membrane protein band 4.1 (Epb41/Protein 4.1)	BS, PCA	(Anderson & Lovrien 1984)	Associated with red blood cell shape and deformability
NC_022018.1_2333634	ENSMOCG00000021685	Mannose receptor C-type 1 (Mrc1)	LFMM, BP	(Moreno-Navarrete et al. 2013; Scott et al. 2015)	Associated with oxidative energy metabolism, expression of Mrc1 is increased in high-altitude deer mice
NC_022028.1_32376384	ENSMOCG00000014534	Mitochondrial calcium uniporter (Mcu)	BP, Pcadapt	(Gherardi et al. 2019)	Associated with mitochondrion metabolic fuel preference
NC_022032.1_3275150	ENSMOCG00000008558	Nuclear receptor binding SET domain protein 3 (Nsd3)	LFMM, BP, BS, PCA	(Ghorbel et al. 2010)	A histone lysine methyltransferase, downregulated in humans with tetralogy of Fallot (chronic hypoxia)
NW_004949096.1_5014291	ENSMOCG00000016806	Leucine rich repeat containing G protein-coupled receptor 4 (Lgr4/Gpr48)	BS, PCA	(Song et al. 2008; Sun et al. 2015)	Linked to fuel adaptation in skeletal muscles, also required during red blood cell production
NW_004949096.1_5682450	ENSMOCG00000021978	Anoctamin 3 (Ano3)	LFMM, BP, BS	(Huang et al. 2013)	Inhibits pain signaling
NW_004949099.1_13906686	ENSMOCG00000020297	Eukaryotic translation initiation factor 4E family member 3 (Eif4e3)	BP, BS	(Ruibin et al. 2018)	Involved in the response to chronic oxidative stress
NW_004949117.1_4902683	ENSMOCG00000005844	Dendrocyte expressed seven transmembrane protein (Dcstamp)	BS, PCA	(Sawatani et al. 2008)	Linked to immunological functions carried out by dendritic cells
NW_004949170.1_267213	ENSMOCG00000019515	PR/SET domain 15 (Prdm15)	BS, PCA	(Mzoughi et al. 2017)	Plays a role in embryonic development and cell fate decision

3.5 Discussion

Despite strong patterns of neutral population structure and IBD across populations, we were able to detect signatures of divergent selection, suggesting adaptive divergence as a result of altitudinal selection across six populations from different altitudes of *M. arvalis*. We did so by genotyping a total of 19,199 SNPs amongst populations and used landscape genomic approaches to identify loci that show patterns of spatially divergent selection. Of the many loci that we identified as outliers, 127 candidate loci were affected by spatially varying selection pressures associated with altitude, and 46 of these are of particular interest as they were located within 37 genes with functional annotation. These provided evidence for functional adaptation to specific selection pressures occurring in high-altitude environments such as low temperatures and hypoxia.

We found high levels of differentiation among study sites and a clear pattern of neutral population structure represented by three major clusters, most likely shaped by a pattern of isolation by distance. High levels of population differentiation are not uncommon for common vole populations. This is caused by recurring local population crashes and limited dispersal capabilities of this species, resulting in low levels of effective gene flow between populations. Our findings are in line with studies on other populations of this species, which found high levels of differentiation on a European wide scale (Heckel et al. 2005; Fischer et al. 2014), as well as between patchily distributed populations no further than 2.5 km apart (Schweizer et al. 2007; Borkowska et al. 2010). Surprisingly, a stream between the two proximate study sites *L1W* and *L2W*, which we expected to form a potent barrier for gene flow (Ratkiewicz & Borkowska 2006), did not result in greater differentiation between these populations compared to differentiation between other proximate populations in this study. Theoretically, only a few migrants each generation are enough to create a panmictic population (Slatkin 1987; Mills & Allendorf 1996) and short distance gene flow is supported by the discovery of descendants of a recent immigrant from *H1* into *H2*, which shows that individuals can travel the distance between these geographically close study sites (Saxenhofer et al. 2019). Thus, it is maybe not surprising that we find a clear signal of population differentiation between distant study sites and only comparatively little differentiation among sites within close proximity. We found that genetic variation was slightly lower and levels of inbreeding and relatedness higher in high-altitude populations compared to low-altitude populations. High-altitude populations might experience increased population fluctuations as a result of longer and colder winters, resulting

in more extreme bottlenecks, which could reduce the amount of genetic variation present (Nei et al. 1975; Frankham 1995). Alternatively, lower levels of genetic variation in high-altitude populations might be a remnant of the recent colonization of higher altitudes in this area (Hamilton et al. 2005; Lischer et al. 2014), and shaped by historical events causing loss of genetic variation due to founder effects and increased genetic drift during range expansion (Heckel et al. 2005; White et al. 2013).

The different genome-scan methods used to detect loci that are potentially involved in local adaptation to high altitude allowed identification of a total of 696 candidate loci. Overlap between different methods was limited, as only 18.2% of outliers were detected by at least 2 methods, and only 7 SNPs were identified as outlier by all four methods. However, such results are not uncommon (de Villemereuil et al. 2014; Harrisson et al. 2017) and are in this case likely caused by the assumed modes of selection that form the basis of each test, which are divergent local selection in the F_{ST}-based methods versus positive selection associated with difference in altitude in the GEA-based methods. Moreover, outlier methods are based on different assumptions regarding demographic effects and use different methods to control for confounding effects of population structure. This results in different outliers detected depending on the methods used (de Villemereuil et al. 2014; Lotterhos & Whitlock 2015). In our analyses, F_{ST}-based methods detected considerable more outliers than GEA-based methods, suggesting that at least some of the outliers detected by F_{ST}-based methods may be related to divergent selection among populations rather than altitudinal selection. The difference between both methods becomes clear when comparing PCA plots based on outliers detected by F_{ST}-based methods with outliers detected using GEA-based methods (Appendix A: Figure S3.5). The PCA plot of loci detected by F_{ST}-based methods shows a pattern very similar to the genome wide pattern of neutral population structure (Appendix A: Figure S3.5a). In contrast, the PCA plot based on outliers detected with GEA-based methods shows a pattern that would be expected from loci shaped by divergence associated with altitude that clearly stand out from the pattern of neutral genetic structure (Appendix A: Figure S3.5b). It is a general problem of genome scans that when model assumptions are not met, increased false positive rates may be the consequence (De Mita et al. 2013; de Villemereuil et al. 2014; Lotterhos & Whitlock 2014). As a conservative approach, we required candidate loci to be detected by at least two genome scan methods. Although it remains possible that the association of certain SNPs with altitude is resulting from drift or allele surfing during range expansion (Klopfstein et al. 2006; Hofer et al. 2009),

geographical proximity and common evolutionary background makes this relatively unlikely in our case (Fischer et al. 2011). Moreover, we identified several candidate genes that have been linked to high altitude adaptation in previous studies.

Organisms living at high altitude are confronted with considerable physiological challenges caused by reduced oxygen levels, low ambient temperatures and increased UV radiation. Therefore, it is perhaps unsurprising that there is ample evidence for adaptive changes as a result of residing at high altitude (reviewed in Cheviron & Brumfield, 2012; Storz, Scott, & Cheviron, 2010)). In humans and other large mammals, these changes involve adaptations in oxygen binding, energy metabolism and responses to UV radiation (Storz et al. 2010; Ai et al. 2014; Foll et al. 2014; Simonson 2015; Wei et al. 2016; Bigham 2016). To ensure proper aerobic performance for thermogenesis, the tissue has to be adequately provided with oxygen, which is impaired in the hypoxic conditions at high altitude (Storz et al. 2007; Storz et al. 2010). High-altitude residents have evolved numerous adaptations in the oxygen delivery system to increase the amount of oxygen transported to the tissue. Interestingly, many of the genes found to be under selection linked to hypoxia in these studies are related to the hypoxia-inducible factor (HIF)-pathway. This pathway includes numerous transcription factors that function as master regulators of oxygen homeostasis by coordinating the transcriptional response to hypoxia (Semenza 2007; Semenza 2009; Bigham & Lee 2014). Cold, hypoxic high-altitude environments are particularly challenging for small endothermic mammals. They only have a small tissue-volume to produce heat compared to their large surface area via which they lose heat. Consequently, small mammals are constantly challenged by high thermogenic demands. Therefore, they have to sustain high levels of aerobic metabolism to support locomotor activity and thermogenesis, which requires a high rate of oxygen transport (Hayes, 1989; Li et al., 2001; Tucker, 1970). This problem has intensively been studied in high-altitude deer mice (*P. maniculatus*), which have adapted to their environment by changes in oxygen transport, mitochondrial oxidative capacity and metabolic substrate sources (reviewed in McClelland & Scott, 2018; Storz, Cheviron, McClelland, & Scott, 2019).

Out of the 127 candidate loci that we identified, 46 loci reside within 37 known genes, and functional annotation of these genes suggests that many of these are associated with oxygen transport and metabolic processes, probably as a response to chronic low oxygen levels and increased metabolic demands at high altitude. We identified signals of selection in two genes

that are involved with the formation of blood vessels. *Arntl* is involved in the regulation of vascular development (Jensen et al. 2012; Jensen & Cao 2013). Likewise, *Map2k5* interacts with the HIF-pathway to control angiogenesis (Doebele et al. 2009; Biyashev et al. 2010) and is found to be under selection in high-altitude Mongolians (Xing et al. 2013). These findings are in line with genetic changes found in high-altitude populations of deer mice that cause greater capillarity in skeletal muscles when compared to their lowland counterparts, and are assumed to have been selected because they should increase the capillary surface area and the capacity to extract oxygen from the blood (Lui et al. 2015; Scott et al. 2015). Other evolved changes in high-altitude populations include a dampened rise in blood hemoglobin (Hb) levels in hypoxic conditions compared to low-altitude populations (Black & Tenney 1980; Beall & Reichsman 1984; Monge & Leon-Velarde 1991). This blunted rise is considered an attenuation of the maladaptive increase in blood Hb levels that increases blood viscosity and impairs oxygen transport (Beall et al. 2010; Storz et al. 2010; Petousi et al. 2014). Additionally, they have evolved increases in the Hb-O_2 affinity of red blood cells to enhance oxygen delivery, resulting from genetically based changes in Hb-functioning (Storz 2016; Natarajan et al. 2018). We identified a signal of selection in *Rora*, which induces transcription of hypoxia-inducible-factor-1alpha (*HIF-1a*) in the HIF-pathway (Kim et al. 2008) and which was found to be strongly associated with phenotypic variation in Hb concentration of Amhara Ethiopians living at different altitudes (Alkorta-Aranburu et al. 2012). We further identified a number of genes that are involved in erythropoiesis. For instance, *Gata1* is a transcriptional regulator of erythropoiesis, which is upregulated under hypoxic conditions by *HIF-1a* (Pevny et al. 1991; Zhang et al. 2012) and *Epb41* plays a critical role in erythrocyte shape as a constitute of the erythrocyte-membrane cytoskeletal network (Anderson & Lovrien 1984). Changes in these genes might play a role in differences in erythrocyte size and shape found in high-altitude populations compared to low-altitude populations (Bullard et al. 1966; Weber 2007; Zhong et al. 2015).

As a response to the cold hypoxic conditions, high-altitude populations have evolved changes in metabolism and substrate use with some evidence for genetic changes in functionally relevant pathways (Cheviron et al. 2014; McClelland & Scott 2018; Storz et al. 2019). These include a greater reliance on carbohydrate use during exercise compared to low-altitude counterparts (Cheviron et al. 2014; Lau et al. 2017) and high rates of fatty acid oxidation to support thermogenesis (Cheviron et al. 2012). We found signals of divergent selection in a number of

genes that function in metabolic processes associated with substrate selection. For instance, *Arntl*, which influences angiogenesis (see above), also interacts with *HIF-1a* under hypoxia to mediate substrate selection according to the time-of-day (Peek et al. 2017) and *Lgr4* regulates the balance between carbohydrate and lipid metabolism in skeletal muscle (Sun et al. 2015). Likewise, a signal of selection on the derived high-altitude SNP was found in *Mcu*, which is involved in calcium uptake in the mitochondria (Baughman et al. 2011), responsible for mitochondrial metabolism and determining cell fate (Shanmughapriya et al. 2015; Kamer & Mootha 2015). *Mcu* is associated with metabolic substrate selection during exercise in skeletal muscle of mice, and a mouse model lacking *Mcu* led to impaired carbohydrate metabolism and enhanced preference for fatty acid oxidative metabolism (Kwong et al. 2018; Gherardi et al. 2019). Finally, we identified a signal of selection in *Mrc1*. Expression of this gene is correlated positively with mitochondrial gene expression in humans (Moreno-Navarrete et al. 2013) and was more highly expressed in high-altitude deer mice compared to their low-altitude counterparts (Scott et al. 2015). Altogether, the selection signals detected for these genes suggest a substantial genetic basis of adaptation to living at high altitude in *M. arvalis* and contribute by enhancing oxygen delivery and thermogenic capacity.

The observation that some of these genes and related pathways have been associated with high altitude in other studies suggests that high-altitude populations of several species have evolved in a similar manner as a response to the unique conditions at high altitude. These results suggest that adaptation to high altitude is on a molecular level at least to some extent predictable, as has been shown for other adaptive processes (Blount, Lenski, & Losos, 2018; Martin & Orgogozo, 2013; Ujvari et al., 2015). However, the physiology of high-altitude adaptation is complex, and our results suggest that while natural selection has acted in parallel on a small set of genes during the process of adaptation to high-altitude environments in different species, we also found a number of candidate genes that have not been detected in previous studies. Thus, on the molecular level, parallel adaptation in different evolutionary lineages may work on a core set of genes detected in all cases, and accessory sets of genes that differ between each evolutionary lineage that becomes adapted to the respective conditions.

Acknowledgements
We thank Lorena Singer, Melanie Hiltbrunner and Anna Walther for assistance in the field and sample collection. We thank Michaela Preick and Alexandra Trinks for help with lab work. For help and assistance with the ddRAD-protocol we are grateful to Michael Lenhard and the people at the Genetics group at the University of Potsdam. We thank Axel Barlow and Stefanie Hartmann for helpful discussion and guidance with analyses. Funding was provided by the University of Potsdam (Core area: Functional Ecology and Evolution). GH was supported by grant 31003A-176209 from the Swiss National Science Foundation.

Data accessibility
Raw read, SNP genotypes, input files for analyses, code and other relevant files will be deposited in the Dryad Digital Repository.

Author contributions
R.F., J.A.E., G.H. and M.H. designed the study. R.F. collected samples. R.F. and J.H. performed laboratory work. R.F. performed the data analysis. R.F. and M.H. wrote the original manuscript. All authors contributed to revisions of the manuscript.

Chapter 4: Article III

Genomic signatures of climate adaptation in bank voles*

Remco Folkertsma, Nathalie Charbonnel, Jana A. Eccard, Heikki Henttonen, Marta Heroldová, Otso Huitu, Petr Kotlík, Emiliano Manzo, Johanna L.A. Paijmans, Attila D. Sándor and Michael Hofreiter

* This manuscript is in preparation for "the Journal of Evolutionary Biology"

4.1 Abstract

Detecting spatially varying selection and adaptive variation can provide insight into a species' ability to adapt to different environments. However, despite advances in genomic technologies, detecting the footprints of selection remains challenging in natural populations. In this study, we analyzed ddRAD sequencing data (21,892 SNPs) to examine population structure and use geographic variation in climate to test for signatures of adaptive differentiation across twelve populations of bank voles (Myodes glareolus) distributed throughout Europe. Accurately detecting signals of selection requires controlling for population structure, which can be challenging when collinearity between climate and population structure exist. We therefore used methods that control for population structure. First, we used two univariate genotype-environment association methods to identify loci under spatially varying selection associated with climate variation. We then employed a multivariate approach, redundancy analysis (RDA), to additionally identify subtle signatures of polygenic adaptation. These methods identified a total of 213 candidate loci, of which 74 were found within genes. These genes were associated with functions related to energy homeostasis and immune system functioning. We then used the result of the RDA to show that climate and population structure have similar effects in shaping genetic variation, but that a part of genetic variation (48.7%) can be accounted for by their joint effects. Thus, making it difficult to separate signatures of local adaptation from neutral patterns of population structure. Further, by examining outlier loci only, we find that among the tested climate variables, annual mean temperature is one of the main factors driving adaptive genetic variation. By using a combination of landscape genomics approaches, our study sheds light on the genome-wide adaptive differentiation and the spatial distribution of variants underpinning adaptive variation shaped by local climates in bank voles.

4.2 Introduction

Understanding how organisms adapt to their local environment is one of the central questions of evolutionary biology. It is generally accepted that genetic variation within and between populations is affected by the local environment in which organisms reside. If individuals are locally adapted, they have higher fitness in their native habitat than individuals from other populations (Kawecki & Ebert 2004; Storz 2005; Hereford 2009). Natural selection acting on phenotypic traits responsible for local adaptation can lead to changes in locally adapted populations phenotypes. Evidence for divergent selection acting across clinal variation in environments is evident from phenotypic differentiation among populations (Stinchcombe et al. 2004; Stillwell 2010). Which is assumed to be driven by a combination of phenotypic plasticity and variation in selection pressures stemming from local environmental conditions acting on genetic variation (Kawecki & Ebert 2004; Savolainen et al. 2013; Diamond & Martin 2016). The genetic basis for environmental adaptation has been uncovered for a few obvious traits with clear phenotypic characteristics, such as variation of coat color in mice associated with environmental background color (Nachman et al. 2003; Linnen et al. 2009), reduction of armor plating in sticklebacks as a response to freshwater colonization (Cresko et al. 2004; Colosimo et al. 2005). These examples aside, information on phenotypic variation is often scarce for many wild species, which makes it difficult to detect the genetic basis of environmental adaptation.

However, positive selection due to local adaptation leaves genomic signatures and genome-scans can be used to pick up these signals that vary strongly within and between populations (Rellstab et al. 2015; Hoban et al. 2016). Genomic scans make use of variation in thousands of SNPs across the genome to identify loci that harbor signals of selection. F_{ST}-outlier tests can be used to detect loci showing signatures of divergent selection (Foll & Gaggiotti 2008; Whitlock & Lotterhos 2015; Luu et al. 2017), but these do not provide information on the agent of selection. In contrast, genotype-environment association (GEA)-tests, test for significant correlations between SNP allele frequencies across populations and variation in environmental variables of interest and thus reveal clues about the important drives of selection (Joost et al. 2007; Coop et al. 2010; Günther & Coop 2013; Frichot & François 2015). Theoretical and empirical studies suggest that the adaptive traits responsible for adaptation to climate are commonly of polygenic nature and are underpinned by many genes of small effect (Pritchard & Di Rienzo 2010; Yeaman 2015; Wellenreuther & Hansson 2016). While genome scans

perform well in picking up signals of adaptive loci with large effect sizes, their ability to detect weak signals of polygenic selection is rather limited (Rellstab et al. 2015; Wellenreuther & Hansson 2016; Forester et al. 2018). This is because these methods only focus on one locus at the time, rather than looking at the combined signal of many individual loci simultaneously. Recently, multivariate methods from the field of community ecology have been introduced to the field of population genomics and used to test for signals of adaptive variation (e.g. Babin, Gagnaire, Pavey, & Bernatchez, 2017; Harrisson et al., 2017; Hoey & Pinsky, 2018; Lasky et al., 2012). These methods are able to analyze many loci and multiple environmental predictors at the same time and are thus better suited to detect signals of polygenic selection that are ignored by GEA-methods (Forester et al. 2018; Capblancq et al. 2018). A benefit of multivariate analyses is that they can also be used to inform about the proportion of the genome that is affected by spatial varying selection and identify important environmental predictors (Lasky et al. 2012; Nadeau et al. 2016; Micheletti et al. 2018). As such, it can inform about the relative importance of neutral and adaptive variation and the relevance of environmental predictors in shaping adaptive differentiation.

Recently a surge of studies has been published that use genome-scan methods to focus on adaptive variation in wild populations (Haasl & Payseur 2016). There is a large body of evidence on rodents and other small mammals showing genetic changes related to variation in climate and other clinal gradients. For example, Phifer-Rixey et al. (2018) found climate associated adaptive divergence in gene regulatory regions and genes related to metabolism and immunity, using *Mus musculus* populations along a latitudinal transect in Eastern North America. Harris & Munshi-south (2017) found signals of selection associated with an urban-rural gradient in white-footed mice (*Peromyscus leucopus*) in genes with metabolic functions. Other studies identified signals of selection in genes by using altitudinal gradients (Waterhouse, Erb, Beever, & Russello, 2018, Folkertsma et al., in prep). However, the specific selective forces driving adaptation as well as the genetic loci affected by these, are much less well understood for larger latitudinal gradients than for shorter ones mentioned above.

The bank vole *Myodes* (formerly known as *Clethrionomys*) *glareolus* (Schreber, 1780) is a small Eurasian forest-dwelling rodent. Its distribution ranges from the Northern latitudes of Scandinavia to the Southern European regions of the Mediterranean peninsulas. *M. glareolus* survived in cryptic glacial refugia during the Last Glacial Maximum (LGM) and subsequently recolonized the European continent (Deffontaine et al. 2005; Kotlík et al. 2006; Colangelo et

al. 2012). This resulted in a complex genetic structure, with eight distinct phylogeographic lineages described based on mitochondrial data (Wójcik et al. 2010; Filipi et al. 2015). It is possible that lineage-specific adaptation to local environmental conditions in refugia affected the potential of post-glacial recolonization of each lineage (Tarnowska et al. 2016; Ledevin et al. 2018), suggesting that adaptive differentiation may have played a role in determining the current distribution of lineages (Kotlík et al. 2014; Strážnická et al. 2018). *M. glareolus* is the rodent reservoir of Puumala virus that causes mild hemorrhagic fever with renal syndrome in humans. It is itself capable of tolerating infection (Bernshtein et al. 1999) but suffers only small negative effects on fitness (Tersago et al. 2012; Kallio et al. 2015). This has fueled studies in the genetic basis of pathogen tolerance and adaptive variation in several genes has been identified that is presumably associated with resistance against Puumala hantavirus (PUUV) infection (Guivier et al. 2010; Guivier et al. 2011; Guivier et al. 2014; Dubois et al. 2017; Rohfritsch et al. 2018). *M. glareolus* experience strong multi-annual fluctuations in population size (Stacy et al. 1997; Tkadlec & Zejda 1998; Zhigalskii 2011), have low dispersal ability (Viitala et al. 1994; Deter et al. 2008) and populations can be highly differentiated even at close geographic proximity (Gerlach & Musolf 2000; Redeker et al. 2006). Populations differ phenotypically and display spatial and temporal variation in characteristics such as initiation of breeding after winter, onset and severity of population fluctuation, body size, age structure and pathogen prevalence (Corbet 1964; Hansson 1985; Aalto et al. 1993; Yoccoz et al. 2001; Behnke et al. 2001; Kloch et al. 2010; Eccard & Ylönen 2011). Within its wide distribution, *M. glareolus* experiences a wide range of environmental conditions, which makes it an excellent species to study climate related adaptive variation.

In this study, we investigated the genomic basis of adaptation to climate in *M. glareolus* by using double digest restriction site associated DNA (ddRAD) sequencing to generate 21,892 SNPS in 276 individuals from 12 populations from across the European continent. These populations experienced a wide range of different climatic conditions which enabled us to look for genomic signatures of positive selection correlating with climate variables related to temperature and patterns of precipitation. We further characterized the association of genetic variation with population structure and climate, and identified important climate variables responsible for adaptive differentiation.

4.3 Methods

4.3.1 Sampling

We sampled individuals from 12 locations across the continental European range of bank voles (Figure 4.1). We prioritized sampling from each phylogeographic lineage (Filipi et al. 2015), to ensure to capture genomic variation present in each. The majority of samples for each population came from a single trapping location. In case individuals were sampled along a transect, we used the center of trapping locations as location. Sampling was performed between 2011 and 2015.

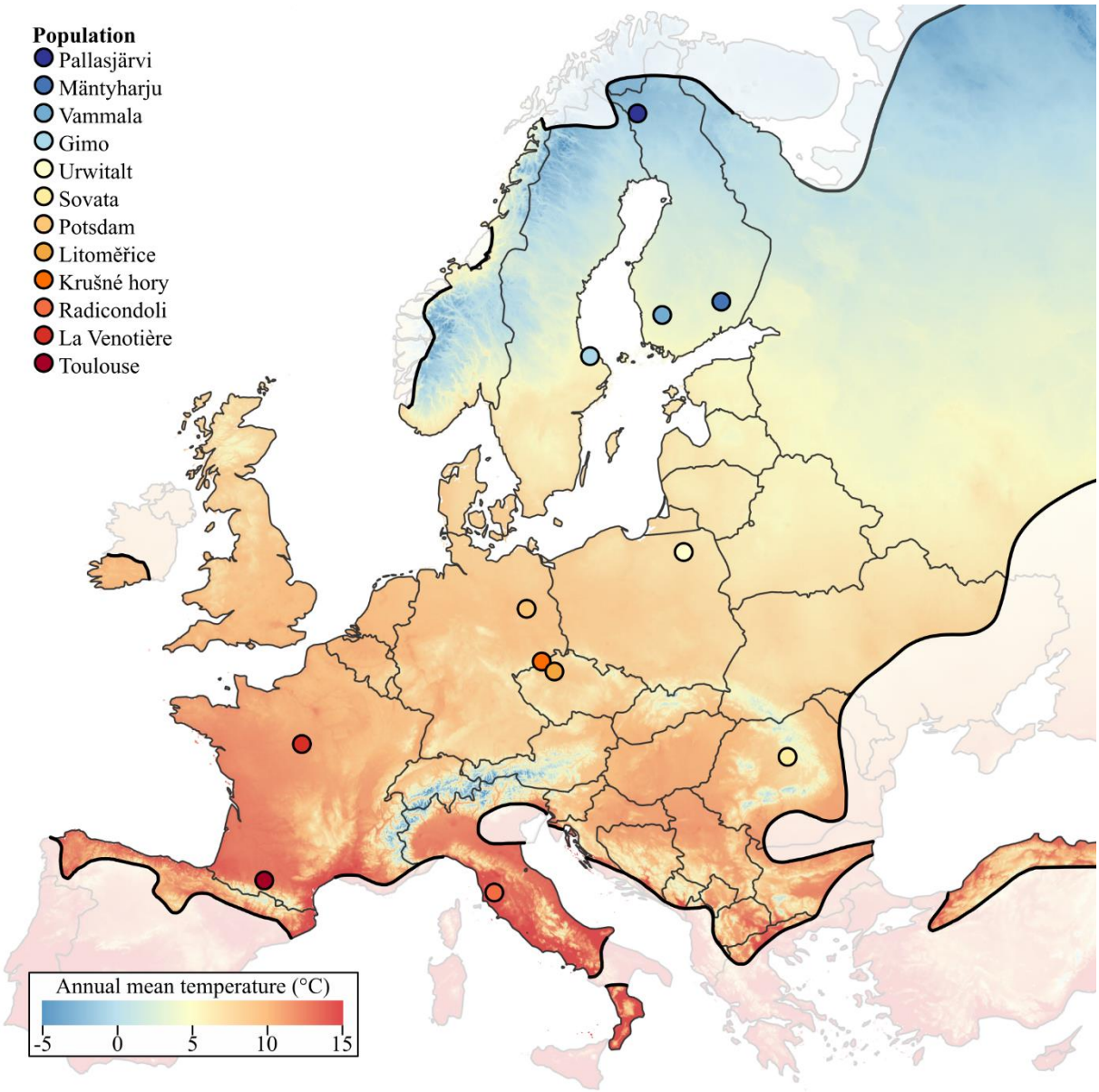


Figure 4.1: Sampling locations of the populations (coloured circles) of *M. glareolus* with annual mean temperature (data: www.worldclim.org (Fick & Hijmans 2017)), distribution: Shenbrot & Krasnov 2005. See table 4.1 for details.

4.3.2 Climatic data

Geographic variation in climate was characterized by downloading a set of 10 bioclimatic variables considered relevant for the biology of bank voles from the WorldClim V2 dataset (http://www.worldclim.org/bioclim), for each of the sampling locations. Climate variables were reduced to principal components to reduce dimensionality using a principal component analysis (PCA) in R v3.4.4 with the prcomp function (R Core Team 2018) (Appendix B: Figure S4.1). This resulted in two climate-based principal components for use in univariate GEA-methods; together these retained 80% of the total original variation. PC1, which explained 62.5% of variation, was associated with variables related to both temperature and precipitation. PC2, which explained 17.1% of variation, associated mostly with precipitation related variables only (Appendix B: Table S4.1).

4.3.3 Molecular methods

Genomic DNA was extracted using the DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's instructions. Extractions were visually inspection for degradation (Graham et al. 2015) on a 1% agarose gel and only non-degraded extracts were selected for further processing. Samples were then analyzed using a double digest restriction associated DNA (ddRAD) sequencing protocol (Peterson et al. 2012). For each sample, 1 ug DNA was digested for 12 hours using EcoRI-HF and MspI-HF restriction enzymes (New England Biolabs, NEB). After double digestion, samples were cleaned using Sera-Mag Speed beads, standardized to 125 ng of digest, ligated with unique in-line barcodes and pooled into pools of maximally 48 individuals before size selection. Pools were size selected using a Pippin Prep system (Sage Science, Beverly, MA) for fragments with 300-400bp insert size providing an estimated 38,000 fragments based on the *M. glareolus* reference genome (GCA_001305785.1) (Lepais & Weir 2014). Finally, size selected pools were PCR amplified to create multiplexed Illumina libraries. Here, we used a cycle number based on the start of the saturation phase on amplification plots, in order to avoid PCR artefacts. The resulting libraries were sequenced on an Illumina NextSeq 500. We first sequenced libraries using 75 bp paired end (PE) sequencing after which we performed another sequencing run using 150 bp PE sequencing.

4.3.4 Sequencing, mapping and genotyping

After sequencing, raw sequences were treated with cutadapt v1.4 (Martin 2011) We remove Illumina specific adapters and trimmed low-quality bases at their 3' end with a phred score

below 20, maintaining only reads with a minimum length of 35bp. These were then demultiplexed using iPyrad v0.7.13 (Eaton 2014). Reads were assigned to individuals according to their barcode not allowing for any mismatch. Subsequently reads were mapped against a modified version of the *M. glareolus* reference genome using bwa-mem v0.7.8 (Li 2013) with default settings. To increase mapping results, we used in silico mate-pair libraries of the Cross-Species scaffolding pipeline (Grau et al. 2018) to increase genome completeness and contiguity of the published *M. glareolus* reference genome. In silico mate-pairs were generated using the Prairie vole (*Microtus ochrogaster*) reference genome (GCA_000317375.1) which diverged about 14 million years ago from *M. glareolus* (Fabre et al. 2012). Running the pipeline increased the N50 by a factor of > 17 and the length of the longest contig by a factor of almost 30, resulting in a 14.2% increase in properly paired reads after mapping (Appendix B: Table S4.2). Samtools v1.6 (Li et al. 2009) was used to convert the generated SAM files into BAM files while filtering for a minimum mapping quality of 30 and only retaining reads mapped in proper pairs. We conducted local realignment around indels using GATK v3.8's (McKenna et al. 2010) RealignerTargetCreator and IndelRealigner modules (DePristo et al. 2011) to avoid erroneous SNP calls due to misaligned reads. Sequencing resulted in notable variation in sequencing depth between individuals, which is not unusual for ddRAD sequencing. As this affects downstream genotyping and population genetic inference (Gautier et al. 2013; Arnold et al. 2013), we excluded 4 samples from our analysis with an average sequencing depth below 3x and applied a down-sampling approach to an average depth of 15x for 11 individuals with a depth substantially above 15x, which resulted in 276 individuals in the final dataset. Finally, as GEA-based methods require called genotypes, genotyping was performed using ANGSD v 0.914 (Nielsen et al. 2012; Korneliussen et al. 2014) with the samtools method (Li 2011) using the BAM files as input. ANGSD is specifically designed for low to moderate coverage data, which comes with greater uncertainty that can lead to biased estimates of allele frequencies and population genetic inferences. Genotypes are inferred using the genotype likelihoods while taking errors in base calling, mapping and alignment into account. Only biallelic SNP sites that passed the following filters were genotyped: a minimum base quality of 20; a minor allele frequency > 0.05 across all samples (minor/major alleles were inferred from the genotype likelihoods); a minimum p-value threshold to call a SNP of 10^{-6}; a minimum depth of 5 and a maximum depth of 100 per sample. Finally, a site had to be present in at least 12 individuals in each of the 12 populations. We further estimated population genetics

parameters and examined population structure based on the likelihood of site frequency spectra (SFS), genotype likelihoods and genotype probabilities. This allowed us to refrain from genotype calling, which can result in bias and errors in downstream population analyses (Johnson & Slatkin 2008; Han et al. 2014). For this, we created a second dataset containing only genotype likelihoods. We used the same filters as described above except that we only included SNPs with a minimum read-depth of 2, we did not allow for missing individuals, and included only one SNP per RAD-tag.

4.3.5 Population genetic inference

Population statistics were calculated using ANGSD, by estimating the site frequency spectrum (SFS), either for each population or population pair (one- and two dimensional SFS) based on genotype likelihoods and polarizing the data by using the reference as ancestral state. Using the one-dimensional SFS, we computed nucleotide diversity as the average number of pairwise differences (π) (Nei & Li 1979) and as the proportion of segregating sites (θ_W) (Watterson 1975), across each scaffold with at least 100 informative sites for each population, which we subsequently used to calculate Tajima's D (Tajima 1989). Accordingly, we calculated the weighted (ratio of sums) pairwise F_{ST} values using the Hudson estimator (Hudson et al. 1992; Bhatia et al. 2013) for each population pair, based on the corresponding two-dimensional SFS. Similarly, heterozygosity was calculated for each individual by estimating the SFS for each sample separately. We estimated per-individual inbreeding coefficients (*F*) with ngsF (Vieira et al. 2013).

To infer population structure and admixture among populations, we used the software NGSadmix (Skotte et al. 2013). NGSadmix is software specifically designed to deal with low coverage sequencing data that uses a maximum-likelihood approach based on genotype likelihoods. In our population genetic analyses, we only selected sites with a minimum read-depth of 2, shared by all individuals, and only included 1 site per RAD-tag, resulting in 2,476 independent sites. We ran NGSadmix with the number of clusters *K* ranging from 2 to 14. For each value of *K*, the analysis was repeated 20 times. We then measured the variance of likelihood values of each set of 20 runs to compare validity across different *K*-values. Here we only report the runs with the highest likelihood solutions for each *K*. We further explored population structure and relationships between individuals by performing a principal component analysis (pca) using the same dataset. By using ngsCovar from the ngsTools suite

(Fumagalli et al. 2014), we approximated the covariance matrix among individuals. We then calculated principal components in R v3.4.4 (R Core Team 2018) using the 'eigen' function and drew pca plots using an in-house R script. We further tested for patterns of isolation by distance (IBD) by testing the correlation between pairwise linearized F_{ST}-values [$F_{ST}/(1-F_{ST})$] and the log-transformed pairwise geographic distance (Rousset 1997), using a Mantel test with 1,000 permutations in the R package 'Vegan 2.5-4' (Dixon 2003).

4.3.6 Genome scan methods

In order to detect signatures of local adaptation to climate under spatially varying selection, we sought for genomic markers that show the strongest relationship between changes in allele frequencies and climatic variation among populations. For this purpose, we used two univariate genotype-environment association (GEA) methods that test for association between allele frequencies and climate, while taking into account neutral population structure.

The first univariate approach, Bayenv2 (Coop et al. 2010; Günther & Coop 2013), corrects for shared demographic history by taking into account the effects of population structure using a neutral covariance matrix of allele frequencies. We used a set of 3,000 non-linked SNPs and 100,000 Markov Chain Monte Carlo (MCMC) iterations to estimate the covariance matrix. We then tested for association of each SNP with the two climate-based principal components, while using the covariance matrix as a null-model to control for demographic history, using 100,000 iterations of the MCMC. To account for run to run variability, we used different seeds for 10 independent Bayenv2 runs and calculated median Bayes Factor (BF) and Spearman's rho (ρ) (Blair et al. 2014). We only considered SNPs as outliers that were among the top 5% of BF and were also in the top 5% of the absolute values of ρ.

We used Latent factor mixed models (LFMM) from the R package LEA v1.6.0 (Frichot & François 2015) as a second univariate approach to test for association between allele frequencies and two climate-based principal components. LFMM corrects for underlying population structure by introducing hidden latent factors that account for neutral genetic structure within the data (Frichot et al. 2013). To determine the appropriate number of latent factors (K), we analyzed the genetic structure in the data based on the best estimates of genetic clusters inferred by sNMF's ancestry estimation program (Frichot et al. 2014). We ran sNMF for each number of ancestral populations (K) from 7 to 15, using alpha = 10 and 20 replicate runs, to determine the cross-entropy-score (CE) for each K. The best value of K was determined

by comparing the average CE-scores values of each *K* and taking the *K* with lowest average score (*K* = 10) (Appendix B: Figure S4.2), from which we chose the run with the lowest CE-score for population assignment. We then performed 10 independent runs of LFMM for both climate-based principal components with *K* = 10, using 50,000 burn-in cycles followed by using 100,000 iterations. From these, we calculated the median z-scores and re-adjusted p-values using the genomic inflation factor (Devlin & Roeder 1999). To correct for multiple testing, a Benjamini and Hochberg FDR correction was applied (Benjamini & Hochberg 1995) and loci with an expected FDR of 10% were identified as outliers. As the results can be sensitive to the number of latent factors included in the model, we also performed the analysis using *K* = 11, which had the second lowest CE-score, and the intuitive value of *K* = 12, for the number of populations we sampled. We found considerable overlap of loci detected as outliers using different values of *K*. As higher values of *K* resulted in a higher number of candidate loci, we decided to only report candidate outlier loci detected using *K* = 10 as a conservative approach.

Next to univariate methods, we also used a multivariate method to test for signatures of adaptation. This method, redundancy analysis (RDA), is an ordination-based approach which extends multiple linear regression to find sets of linear combinations between a set of multiple response variables (SNP allele frequencies) and multiple predictors (climate and population structure) (Legendre & Legendre 2012). As RDA takes advantage of covarying signals of selection across loci, this method is able to detect signals of weak, polygenic selection (Forester et al. 2018). RDA and associated analyses were performed using the Vegan 2.5-4-package in R (Dixon 2003). The dependent matrix contained SNP allele frequencies for each population. As an independent matrix we first used the ten climatic variables (standardized to zero mean and unit variance) as predictors, but this resulted in over-parametrization of the model and multicollinearity among predictors. We therefore removed variables based on their variance inflation factor (VIF), starting with the highest, until all predictors had VIF < 10 (Zuur et al. 2010). This resulted in 5 remaining climatic variables (Mean diurnal range, annual mean temperature, temperature seasonality, annual precipitation and precipitation seasonality). To control for the effects of population structure we performed a partial RDA (pRDA) (Lasky et al. 2012; Harrisson et al. 2017). Here, we used the climatic variables as constraining variables and, to correct for population structure, the values of each population on the first 4 axis of the ngsCovar analysis as conditioning variables. We identified SNPs as potential outliers that were in the tails of the distribution for each of the first two constrained axis, by using a cut-off that

was ±3 SD from the mean score, as this renders the best balance between true- and false-positive rates (Forester et al. 2018).

4.3.7 Variance partitioning and the identification of important climate variables

RDA can also be used to evaluate the variation that can be explained by the constraints used. For this purpose, we used RDA and pRDA to estimate the proportion of genetic variance explained by climate, population structure and their joint effects (describing spatially structured climatic variation), using the variance values (inertia) in the constrained matrix of the respective partial models. Significance of these models and the marginal effects of each climate variable were tested using Anovas with 1,000 permutations. To further assess the independent effects of each climate variable and to calculate the amount of genetic variance each explains, we performed multiple pRDA. In these, we conditioned the effect of each climate variable on the others and thus removed the variance due to partial correlations with other variables. In addition, to identify the climate variables that contributed most to adaptive variation, we performed pRDA on a subset of loci identified as outliers. Here, we assessed the amount of genetic variance in outlier loci that can be explained by climate variables. The significance of these models was tested using Anovas performed with 1,000 permutations. We further identified the variable that had the largest influence on each SNP by ranking marker scores for each pRDA for each climate variable. For each SNP we identified the climate variable with the highest rank (after Babin et al., 2017).

4.3.8 SNP annotation and gene ontology

As a well annotated high-quality genome for *M. glareolus* is missing, we attempted to find homologous *M. ochrogaster* positions for each candidate SNP using the LastZ pairwise alignment tool v1.04.00 (Harris, 2007). For this, we selected 20,000 bp scaffolds surrounding each outlier SNP and LastZ default options to compute pairwise alignments. Only alignments with a bitscore above 1,000 were kept from which we selected the longest alignments. These were used to determine the homologous position of each candidate SNP on the *M. ochrogaster* genome. We then used biomart (Smedley et al. 2015) and the Ensembl genome database to retrieve associated gene ontology (GO) terms of the *M. musculus* orthologs using the Uniprot knowledgebase (Bateman et al. 2017). CateGOrizer was used to analyze biological process GO-categories and count GO-terms using the GO-slim method (Hu et al. 2008).

4.4 Results

4.4.1 Sequencing results

We used a ddRAD-sequencing approach for genome-wide genotyping, resulting in a dataset of 276 *M. glareolus* from 12 locations across the Europe continent. In total we obtained 613 million reads during two runs of sequencing. After filtering for low quality reads, assigning individuals to barcodes, removing low coverage individuals and downsampling high coverage individuals, this resulted in a dataset with an average of 1,615,845 (SD = 747,121) high-quality properly paired PE reads per individual that aligned to our improved reference genome. Sequencing results differed among populations, with per population averages ranging between 964,915 (Vammala) and 2,608,954 (Potsdam) PE reads. High quality called genotypes covered an average of 8,477,465 (SD = 2,145,583) nucleotides of the genome (~0.33%) per individual. This resulted in a dataset with an average sequencing depth of 15.94X (SD = 5.41), ranging from 7.31X to 29.41X per individual. Using this data, we retrieved information for 21,892 SNPs on 7,679 RAD-tags shared between 12 individuals in each population, that can be used in genotype-environment analyses. The dataset with genotype likelihoods only for population genetic inference consisted of data for 2,476 variable sites with data for all individuals.

4.4.2 Population inferences

We used the one-dimensional SFS to estimate genetic diversity statistics for each of the 12 *M. glareolus* populations (Table 4.1). The proportion of segregating sites and the average number of pairwise differences differed slightly between populations. The median value for θ_W was 0.0027 and varied approximately threefold ranging from 0.00160 (Gimo) to 0.00483 (Sovata), while the median value for π was 0.00316 and varied approximately two-fold between 0.00188 (Gimo) and 0.00407 (Radicondoli). Genome-wide average Tajima's D was highest for Toulouse (0.584) and positive for nine populations, while Sovata had a negative (-0.671) Tajima's D, just as three other populations. Positive Tajima's D values often indicate a recent population contraction, while negative Tajima's D values are often an indication for a population expansion after a recent bottleneck. Heterozygosity was generally high across populations, but differed significantly between populations, ranging from 0.0012 (Gimo) to 0.0025 (Sovata). Surprisingly, Sovata, which has the highest level of heterozygosity displays the largest level of inbreeding, but, although this seems counter-intuitive, inbreeding coefficient and

heterozygosity do not necessarily correlate (Slate 2004). Finally, kinship coefficients (*r*) and inbreeding coefficients (*F*) varied among populations, but were overall low.

Table 4.1: Overview of of characteristics and genetic diversity statistics (mean ± standard deviation) of 12 sampled *M. glareolus* populations. Sample size (n), Watterson's theta (θ_W), Tajima's pi (π), Tajima's D, heterozygosity - the proportion of heterozygous genotypes, average population inbreeding coefficients (F_{IS}) and the average population kinship coefficient (*r*) are displayed.

Population	ID	Latitude	Longitude	n	θ_W	π	Tajima's D	Heterozygosity	F_{IS}	r
Pallasj rvi	Pal	68.011 N	24.140 E	24	0.0026 (0.0016)	0.0029 (0.0020)	0.326 (1.06)	0.0019 (0.00007)	0.014 (0.030)	0.004 (0.024)
M ntyharju	Man	61.479 N	26.870 E	24	0.0025 (0.0015)	0.0028 (0.0021)	0.293 (1.05)	0.0018 (0.00006)	0.004 (0.011)	0.010 (0.039)
Vammala	Vam	61.377 N	22.822 E	23	0.0024 (0.0015)	0.0026 (0.0021)	0.244 (1.08)	0.0016 (0.00013)	0.007 (0.014)	0.006 (0.030)
Gimo	Gim	60.563 N	17.822 E	24	0.0016 (0.0012)	0.0019 (0.0018)	0.318 (1.15)	0.0012 (0.00010)	0.003 (0.007)	0.009 (0.035)
Urwitalt	Urw	53.800 N	21.650 E	22	0.0028 (0.0015)	0.0032 (0.0020)	0.453 (0.93)	0.0021 (0.00003)	0.016 (0.025)	0.007 (0.030)
Potsdam	Pot	52.436 N	13.041 E	24	0.0026 (0.0014)	0.0031 (0.0019)	0.569 (0.94)	0.0020 (0.00007)	0.025 (0.034)	0.002 (0.014)
Krušné hory	Kru	50.676 N	13.561 E	24	0.0035 (0.0019)	0.0034 (0.0023)	-0.186 (0.90)	0.0021 (0.00015)	0.012 (0.026)	0.005 (0.032)
Litoměřice	Lit	50.540 N	13.937 E	23	0.0033 (0.0017)	0.0034 (0.0021)	-0.011 (0.88)	0.0021 (0.00008)	0.018 (0.025)	0.005 (0.030)
Sovata	Sov	46.623 N	25.114 E	23	0.0048 (0.0021)	0.0040 (0.0023)	-0.671 (0.76)	0.0025 (0.00016)	0.028 (0.023)	0.005 (0.023)
Radicondoli	Rad	43.259 N	11.092 E	22	0.0035 (0.0018)	0.0041 (0.0024)	0.518 (0.85)	0.0025 (0.00016)	0.012 (0.029)	0.005 (0.042)
La Venoti re	Ven	47.746 N	1.775 E	21	0.0029 (0.0014)	0.0033 (0.0020)	0.444 (0.93)	0.0021 (0.00011)	0.006 (0.012)	0.008 (0.035)
Toulouse	Tou	43.237 N	0.824 E	22	0.0020 (0.0013)	0.0025 (0.0020)	0.586 (1.06)	0.0016 (0.00013)	0.003 (0.011)	0.008 (0.036)

4.4.3 Population structure

We found a clear pattern of distinct and differentiated *M. glareolus* populations across the European continent. Results from the principal component analysis showed a clear structuring of individuals in populations along the first 4 principal components (Figure 4.2), which together explained 20.14% of the total genetic variation. Component 1 explained 9.42% of genetic variation and separated populations along a latitudinal gradient, with a cluster of individuals from South-western populations, a second cluster with individuals from the Central European populations and Gimo in Sweden and a third cluster including all remaining Scandinavian populations. The second component explained 5.17% of genetic variation. It separated Gimo from the other Scandinavian populations and separated the three geographically close populations in Central Europe from the Eastern European populations. The third and fourth components both explain less than 5% of genetic variation, and these separate Gimo and Radicondoli from the other populations. Interestingly, none of the first four components separated the Finish populations from each other.

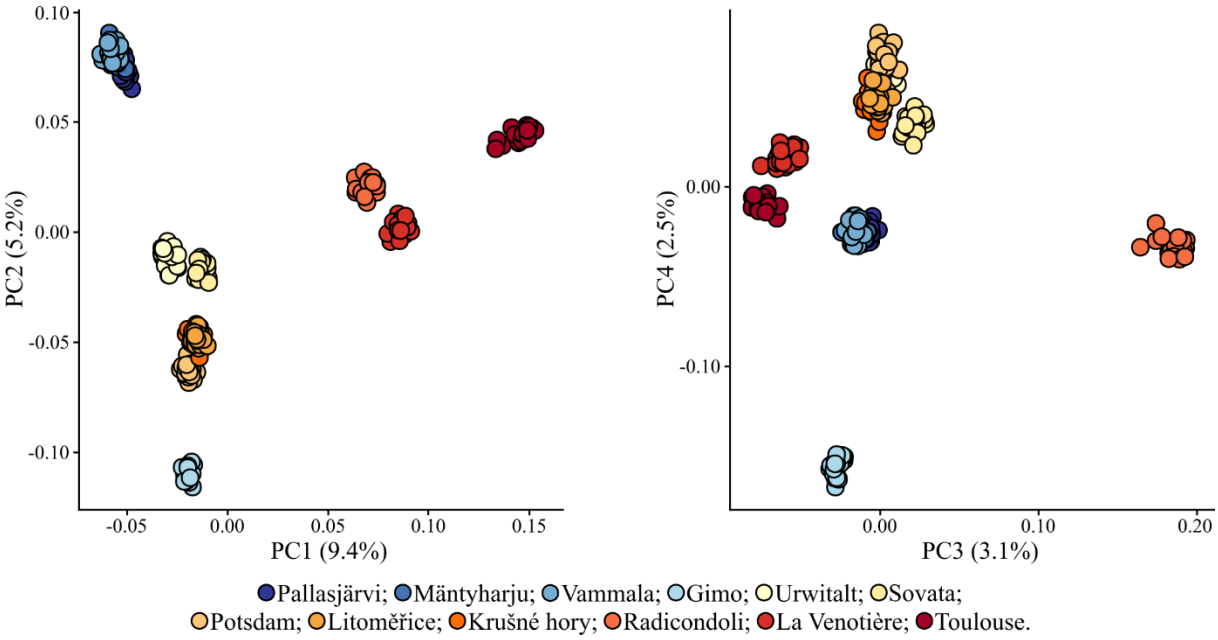


Figure 4.2: Genetic structure of populations as depicted by the first four principal components of a PCA based on pairwise genetic covariance among 2,476 variable sites of 276 individuals from 12 *M. glareolus* populations, together components explain 20.1% of genetic variation.

Admixture analysis using different values of K showed a consistent pattern of differentiated populations (Figure 4.3). The lowest variance was found for 2 or 3 ancestral populations, but higher K-values reveal additional sub-structuring related to geographical location. The 2nd lowest variance was found for 11 ancestral populations. By using this *K* value, individuals are grouped into distinct clusters based on sampling locations, except for the individuals from Mäntyharju and Vammala, which form a single cluster. When 12 ancestral populations are assumed, individuals are grouped into clusters based on population locations, with only some degree of admixture suggested between individuals from Mäntyharju and Vammala, as well as between individuals from Krušné hory and Litoměřice. Decreasing the number of ancestral populations down to $K = 5$, suggests similarity of geographically close populations as individuals from geographically close areas start to cluster together, which coarsely resembles clustering of individuals in the principal component analysis. A similar pattern emerges from the results of the genomic pairwise F_{ST}-values, which revealed moderate to high levels of differentiation between populations (Appendix B: Table S4.3). Here, pairwise F_{STS} corresponded well with geographic proximity of populations. This is supported by Mantel tests, as genetic distance and geographic distance were highly correlated ($r = 0.47$, $p = 0.002$), suggesting a pattern of IBD. Interestingly, the population from Gimo is more differentiated from the Scandinavia populations than it is from Central European populations.

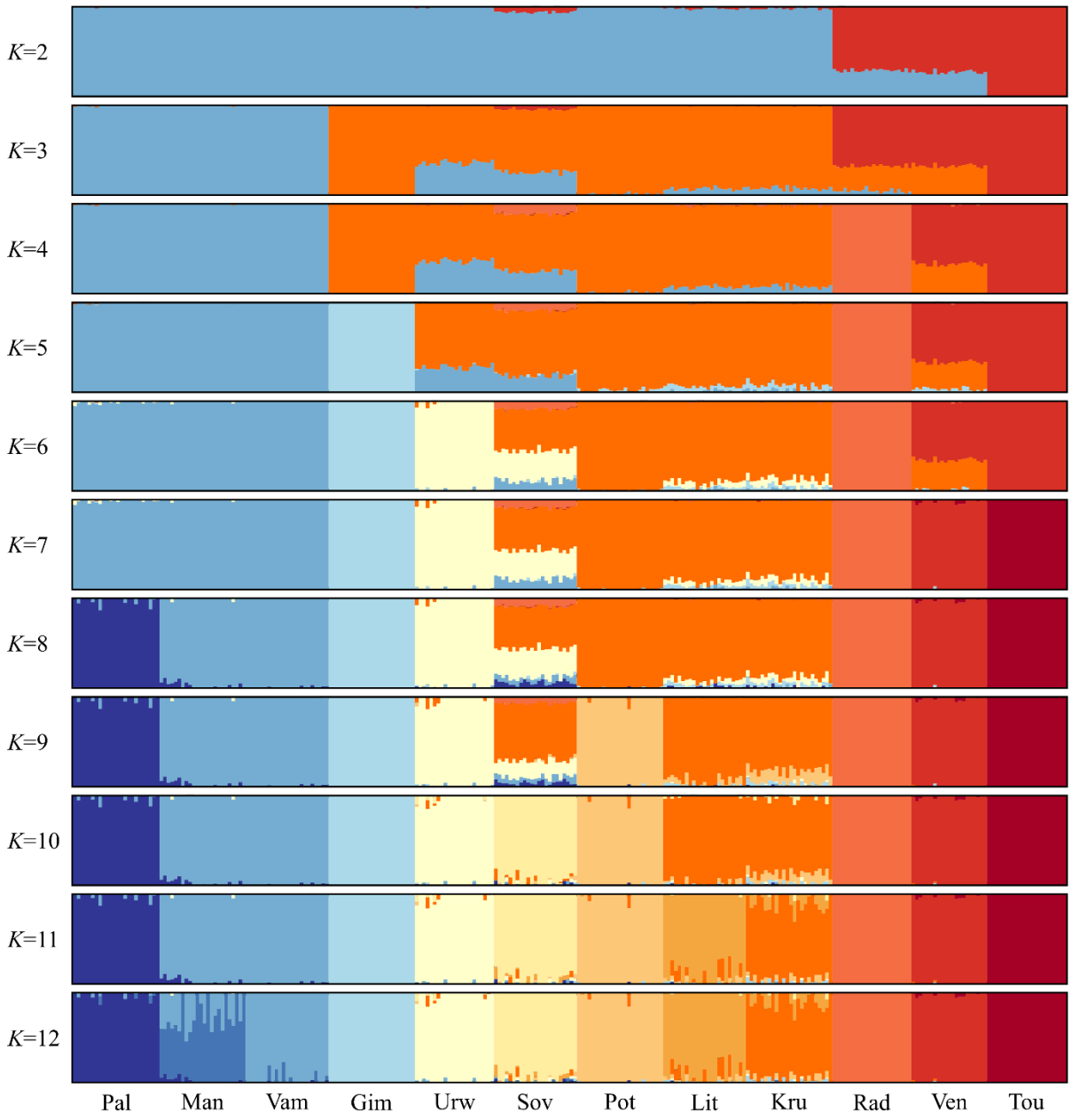


Figure 4.3: Admixture proportions using NgsAdmix based on 2,476 variable sites of 276 individuals, with various numbers of ancestral populations (*K*= 2-12) of *M. glareolus*. Each individual is represented by a column with colours corresponding to the proportions of their ancestry components. Populations: Pal - Pallasjärvi; Man - Mäntyharju; Vam - Vammala; Gim - Gimo; Urw - Urwitalt; Sov - Sovata; Pot - Potsdam; Lit - Litoměřice; Kru - Krušné hory; Rad - Radicondoli; Ven - La Venotière; Tou -Toulouse.

4.4.4 Candidate loci and Gene ontology

Univariate approaches detected a total of 975 outlier loci which were found to be associated with climate. These were located on 730 individual ddRAD-tags. Analyses using Bayenv2 detected 631 outlier loci, among which 283 loci had a strong association with PC1 and 354 loci were associated with PC2. LFMM was more conservative and detected a total of 497 outlier

loci, among which 134 loci were associated with PC1 and 377 with PC2. Overlap between methods was low with 152 outlier loci discovered by LFMM as well as Bayenv2 (15.6% of total). The multivariate RDA detected 485 outlier loci associated with the first 2 RDA axis. Overlap between univariate approaches and RDA was poor, with only 69 loci detected by both approaches (5.1%). Overall, a total of 1,392 outlier loci were detected using all methods (Appendix B: Figure S4.3), of these, we only considered outlier loci detected by at least 2 methods as plausible candidates, resulting in 213 candidate loci. LastZ pairwise alignment of 20kb scaffolds surrounding these resulted in 209 scaffolds that aligned to the *M. ochrogaster* genome, from which we determined the homologous SNP position. Of these, we found that 8 candidate loci were found in exons, 86 in introns and 115 in intergenic regions, resulting in 74 genes with one or more candidate loci (Appendix B: Table S4.4). Several of these genes were associated with functions in metabolism, immunology and DNA-break repair (Table 4.2). The GO-terms based on the *M. musculus* annotation most commonly assigned to genes were metabolism, cell communication, development and signal transduction.

4.4.5 Variance partitioning and the identification of important climate variables

The RDA model with only climate variables (C) revealed that a significant amount of genetic variance in all SNPs could be explained by climate ($F = 2.09$, $p = 0.005$), which explained 33.2% of genetic variance (adjusted R^2) among populations. In this model, only the marginal effects of mean diurnal range and temperature seasonality were significant and these also had the largest independent effects both explaining about 12% of variance while controlling for other climate variables. When conditioning climate on population structure (C | S), the model lost significance ($F = 1.34$, $p > 0.2$), but the constraining factors (i.e. climate variables) still explained 6.9% of the total genetic variance. However, none of the marginal effects of the climate variables was significant, and each independent climate variable explained less than 5% of the total variance (Figure 4.4, Table 4.3). Climate and population structure combined (C + S) explained 71.4% of genetic variance ($F = 4.06$, $p = 0.001$). Using pRDA, we identified that population structure explained a larger part of the total variance explained than climate. While controlling for population structure (C | S), climate explained 18.4% of the total genetic variance, and vice-versa, population structure explained 33.0% of genetic variance while controlling for climate (S | C). The joint effect of population structure and climate was large and explained 48.7% of genetic variance (Table 4.4).

Table 4.2: Subset of top candidate loci identified with at least two genome scan methods. Locus refers to the location on the *M. ochgrogaster* reference genome (Chromosome_base-position). Method of outlier detection refers to the genome scan method that identified the SNP as an outlier and which of the principal components it was associated with.

Locus	Gene (abbreviation)	GeneID (UID)	Method of outlier detection	References	Putative function and relevance
NC_022011.1_19197181	Zinc finger homeobox 3 (Zfhx3)	102001415	Lfmm_PC1, Lfmm_PC2, Bayenv_PC1	(Balzani et al. 2016)	Transcription factor expressed in the suprachiasmatic nuclues with a role in circadian rhythms
NC_022013.1_77189997	Basic leucine zipper ATF-like transcription factor 3 (Batf3)	101980506	Lfmm_PC1, Bayenv_PC1, Bayenv_PC2, RDA	(Lee et al. 2017)	Transcription factor involved in the differentiation of regulatory T cells
NC_022013.1_78476899	Potassium voltage-gated channel subfamily H member 1 (Kcnh1)	101985787	Lfmm_PC2, Bayenv_PC2	(Zhang et al. 2014)	Involved in adipogenic differentiation and production
NC_022017.1_17053779	ADAM metallopeptidase with thrombospondin type 1 motif 20 (Adamts20)	101980054	Lfmm_PC2, Bayenv_PC2	(Silver et al. 2008)	Required for melanoblast survival, which is responsible for coat colour variation in a range of species
NC_022018.1_59867563	Neurotrophic receptor tyrosine kinase 2 (Ntrk2)	101991185	Lfmm_PC2, Bayenv_PC2, RDA	(Kernie 2000)	Linked to regulation of food intake and body weight in rodents.
NC_022024.1_33246556	Insulin like growth factor 1 (Igf1)	101987701	Lfmm_PC2, Bayenv_PC2	(Kroonsberg et al. 1989)	Involved in mediating growth and development
NC_022027.1_54997482	Leucine rich repeat containing 8 VRAC subunit C (Lrrc8c)	101997182	Lfmm_PC2, RDA	(Tominaga et al. 2004)	Associated with early stage adipocyte differentiation
NC_022028.1_49631929	Dynein axonemal heavy chain 8 (Dnah8)	101988652	Lfmm_PC2, Bayenv_PC2	(Söhle et al. 2012)	Axonemal dynein influencing lipid metabolism, possibly by regulation of inflammatory processes
NC_022031.1_30606927	BTB domain and CNC homolog 2 (Bach2)	101995098	Lfmm_PC1, Lfmm_PC2, Bayenv_PC2, RDA	(Roychoudhuri et al. 2013)	Transcription factor involved in controlling the balance between tolerance and immunity
NW_004949096.1_35517029	Biliverdin reductase A (Blvra)	101996347	Lfmm_PC2, Bayenv_PC2	(Baranano et al. 2002)	Facilitates conversion of biliverdin to bilirubin protecting against cell damage
NW_004949099.1_1813079	Aprataxin and PNKP like factor (Aplf)	101983646	Lfmm_PC1, Bayenv_PC1	(Grundy et al. 2013)	Involved in double-strand DNA break repair
NW_004949106.1_3445183	Phospholipase C like 1 (Plcl1)	101997890	Bayenv_PC1, RDA	(Oue et al. 2016)	Involved in adipose metabolism and energy expenditure and is a regulater of non-shivering thermogenesis
NW_004949106.1_7957839	Signal transducer and activator of transcription 4 (Stat4)	101994625	Lfmm_PC2, Bayenv_PC2	(Kaplan 2005)	Plays a major role in the immune system response to viral infections

The RDA (including all climate variables conditioned on population structure) on the subset of 485 outlier loci detected by the RDA analysis, explained 51.1% of variance in these outliers (F = 3.73, P = 0.03). Results from pRDA show that only annual mean temperature explained a significant amount of outlier variance (52.0%, $p < 0.05$). Followed by non-significant effects of mean diurnal range (35.1%, $p = 0.08$) and precipitation seasonality (29.1%, $p = 0.09$) (Table 4.3). The climate variable with the highest influence on outlier SNPs as detected by RDA was annual mean temperature, which was associated with 37.1% of the markers, followed by mean diurnal range which was associated with 20.8% of markers. Of the outlier loci detected by all methods, the primary associations were with annual mean temperature (27.5%) and annual precipitation (25.4%). The same picture arises when only considering candidate loci that were detected by at least 2 methods. Here, 28.6% of loci are associated with annual mean temperature and 25.4% are associated with annual precipitation (Appendix B: Table S4.5).

Table 4.3: Results of pRDA to test for the unique contribution (in terms of amount of genetic variance explained) of each climate variables. Analysis are done separately for all 21,892 SNPs and the 485 outlier SNPs identified by the RDA analysis.

	All 21,892 SNPs			485 RDA outlier SNPs		
Variable	r^2	r^2_{adjust}	P-value	r^2	r^2_{adjust}	P-value
AnMTemp	0.04	0.06	0.28	0.17	0.52	0.03
TempSeas	0.03	0.01	0.43	0.04	0.07	0.23
MDR	0.04	0.06	0.27	0.12	0.35	0.07
AnPrec	0.04	0.05	0.29	0.08	0.20	0.13
PrecSeas	0.04	0.04	0.29	0.10	0.29	0.09

Table 4.4: Results of variance partitioning with the variance (inertia) and the percentage of variance explained by climate only, population structure only, and the joint effect of climate and population structure.

Effect	Variance explained	Percentage explained
Total variance (explained+unexplained)	986.0	
Total explained variance (climate + population structure + climate|population structure)	934.8	94.8
Climate	171.6	18.4
Population structure	308.3	33.0
Joint effect of climate and population structure	454.9	48.7

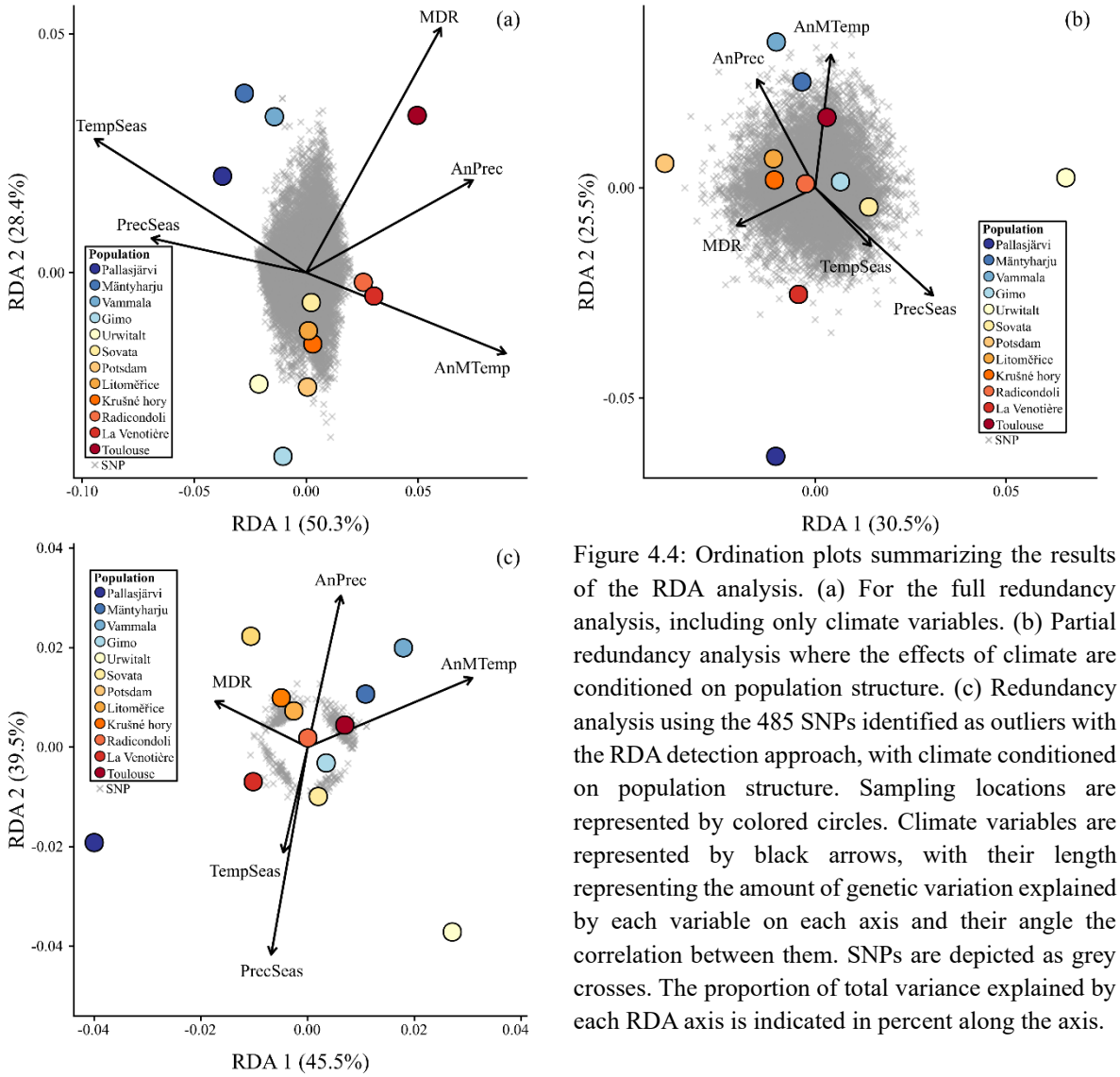


Figure 4.4: Ordination plots summarizing the results of the RDA analysis. (a) For the full redundancy analysis, including only climate variables. (b) Partial redundancy analysis where the effects of climate are conditioned on population structure. (c) Redundancy analysis using the 485 SNPs identified as outliers with the RDA detection approach, with climate conditioned on population structure. Sampling locations are represented by colored circles. Climate variables are represented by black arrows, with their length representing the amount of genetic variation explained by each variable on each axis and their angle the correlation between them. SNPs are depicted as grey crosses. The proportion of total variance explained by each RDA axis is indicated in percent along the axis.

4.5 Discussion

The goal of the present study was to investigate genome-wide patterns of selection associated with climate amongst 12 populations from across the European distribution of *M. glareolus*. To this end, we used a ddRAD-sequencing approach to genotype 21,892 SNPs. By combining approaches to reduce error rate, we discovered numerous correlations between population allele frequencies and environmental variation. In addition to this, we used a multivariate framework and identified distinct climate variables responsible for adaptive genetic variation. By doing so, we were able to gather a picture of genome-wide adaptive variation and identified functional significant genes which play a role in environmental local adaptation.

4.5.1 Genetic diversity and population structure

Patterns of genome-wide genetic variation varied among populations, this is similar to what is observed in other studies using genome-wide markers but on different geographical scales (White et al. 2013; Rohfritsch et al. 2018). Differences in genetic variation among populations might be caused by the fact that not all population across the distribution are cyclic (Bondrup-Nielsen & Ims 1986; Brommer et al. 2010). Moreover, those that are may not be in the same phase of the cycle (Hansson & Henttonen 1985), all of which could affect population statistics measurements (Rikalainen et al. 2012; Norén & Angerbjörn 2014; Dubois et al. 2017).

Populations were highly differentiated and we found a strong pattern of population structure in line with a signal of isolation by distance. These results are not surprising, populations experience recurring population crashes and effective gene flow between populations is low, which results in a strong pattern of isolation by distance across smaller and larger geographic scales (Aars et al. 1998; Gerlach & Musolf 2000; Redeker et al. 2006; Guivier et al. 2011). The phylogeography of *M. glareolus* is marked by distinct mitochondrial (mtDNA) lineages, which resulted from survival within glacial refugia and recolonization of the European continent at the end of the last glaciation (Deffontaine et al. 2005; Kotlík et al. 2006; Filipi et al. 2015). The patterns of population structure found here are broadly comparable with previous results based on mtDNA data, but some incongruence is apparent. For instance, individuals from Pallasjärvi were expected to form a single cluster, as individuals from this population are the only individuals expected to have hybridized with the red-backed vole (*Myodes rutilus)* (Tegelström 1987; Boratyński et al. 2011). However, based on our nuclear data, they consistently form a cluster with other Finish individuals in the PCA, and only separate at higher values of *K* in the admixture analysis. Suggesting, that hybridization at the genomic level is much less apparent. In addition, individuals from Venotière as well as individuals from Sovata do not show any clustering with other individuals from their respective mtDNA clade. Such conflicting geographic patterns between mitochondrial and nuclear genetic markers (mito-nuclear discordance) are not uncommon (Toews & Brelsford 2012) and might be a result of selection on the mitochondrial genome (Boratyński et al. 2011; Boratyński et al. 2016; Bonnet et al. 2017; Stojak et al. 2019). However, as the distribution is broadly defined and clades sometimes overlap (Wójcik et al. 2010; Drewes et al. 2017), these results have to be taken with some caution.

4.5.2 Variance partitioning

Results from the RDA provide the same picture. We are able to explain a large amount of the total genetic variation present among populations. Population structure accounted for a large part of the explained genetic variation, even when taking climate into account. This is in contrast with the effects of climate, which are non-significant and weak when controlling for population structure. Additionally, the contribution of individual climate variables seems to be small, as none of them were significantly associated with genetic variation. Population structure affects the genome as a whole, in contrast, despite the large geographical range of *M. glareolus* and the wide differences in several climate variables at opposite ends of its distribution, the effects of climate should be much more subtle. Affecting only small parts of the genome, specifically those that are responsible for adaptive variation.

4.5.3 Genome scan method

In this study we used different genome scan methods to identify loci that may be involved in local adaptation of *M. glareolus* on the European continent. Despite the strong patterns of population structure, univariate and multivariate methods still were able to identify a total of 1,392 candidate SNPs that show signals of selection associated with climatic variation. However, SNPs identified as outliers differed between methods and overlap between methods was limited. A total of 214 SNPs were identified by at least two methods and 17 candidate SNPs were identified by all three methods. It is not surprising that we find little overlap between methods. Outlier methods are based on different assumptions regarding demographic effects and use different methods to control for confounding effects of population structure (de Villemereuil et al. 2014; Lotterhos & Whitlock 2015; Hoban et al. 2016). First, we used 2 univariate methods, LFMM and Bayenv2, overlap between these univariate methods was low, as only 15.6% of outliers detected by LFMM were also found with Bayenv2. Next to these, we used RDA as a multivariate approach, to complement these analyses. RDA might be better suited to pick up weak signals of polygenic selection and might in addition, be more robust to our random sampling design that does not maximize environmental differentiation (Forester et al. 2018). Indeed, overlap between univariate methods and RDA was low, with only 5.1% of loci detected by RDA also being detected by either LFMM or Bayenv2. Little overlap between different methods presents a strong argument for the use of different approaches when testing for signals of local adaptation.

4.5.4 Identification of important climate variables for adaptive variation

We analyzed candidate loci with RDA to identify climate variables that are most important in shaping adaptive variation in *M. glareolus*' distribution. In general, more loci were associated with among population variation in annual climate than with among population seasonal variation in climate. Annual mean temperature explained the highest proportion of genetic variation of all climate variables tested. Moreover, across all variables and outlier subsets, annual mean temperature was consistently associated with the highest amount of adaptive variation, explaining 28.6% of adaptive variation in the candidate loci subset. Annual mean temperature differed between -0.8C (Pallasjärvi) and 12.4C (Radicondoli) among our populations and can thus be expected to act as an important selective pressure. Although we focus on the importance of annual mean temperature, this variable was highly correlated with monthly and quarterly minimum and maximum temperatures (Appendix B: Table S4.6). However, these variables could not be included in the RDA as this would have over-parameterized the model, but these factors are also of major importance in shaping adaptive variation (David et al. 2005; Challinor et al. 2007; Tarnowska et al. 2016; Stojak et al. 2019). Especially for small mammals, which have limited capacity for insulation and experience increased heat loss compared to larger bodied individuals, colder conditions can be challenging (McClelland & Scott 2018). Consequently, it is expected that individuals from populations living in colder environments experience higher thermogenic demands and have accordingly developed a range of adaptations (Nilsson & Nilsson 2016). Contrary, warmer temperatures might also be difficult to cope with for small mammals, as this can easily lead to overheating (Rezende et al. 2004), which could also lead to specific adaptations in populations in warmer climate. Interestingly, experimental selection for increased aerobic capacity in *M. glareolus* resulted in the evolution of an increased thermogenic capacity, resulting in increased cold tolerance (Stawski et al. 2017). It is thus very likely that temperature shapes a large part of adaptive genetic variation across *M. glareolus*' distribution. However, geographical variation in metabolic traits of is rarely studied in this species. Boratyński et al. (2011) showed that voles at higher latitudes appear to have a lower metabolic rate, which was partly confounded by mtDNA introgression. Another found no variation in BMR of voles from a wide range, which was attributed to adaptation to microclimatic conditions (Aalto et al. 1993). Unfortunately, this study used questionable methodology (different aged individuals among sites, variable housing conditions and variable moments of measurement) and it is likely that results thus do not

represent adaptation to local conditions or even genetic adaptation. Nevertheless, evidence for the importance of temperature in shaping genetic variation comes from studies examining the influence of climate on genetic structure. For instance, the average minimum temperature in January was an important factor in shaping the genetic structure across the landscape in two related vole species (Stojak et al. 2019), while the presence of *M. glareolus* belonging to the Carpathian lineage in a secondary contact zone with the Eastern lineage was positively correlated with mean July temperature (Tarnowska et al. 2016). This suggests adaptation to temperature does take place. Moreover, 7 out of 9 variables identified to be important in shaping the distribution of *M. glareolus* in western Europe were related to indices of environmental temperature (Amirpour Haredasht et al. 2013).

The second most important variable was mean annual precipitation, which explained about 27.8% of adaptive variation in the candidate loci subset. Patterns of precipitation are related to biomass production and plant community composition (Yan et al. 2015). Therefore, variation in precipitation might have an indirect effect on fitness and adaptive variation, by affecting habitat quality and abundance of food, or changes in resource abundance related to water availability. Indeed, *M. glareolus* population cycles and abundance are related to resource availability (Pucek et al. 1993; Imholt et al. 2015) and precipitation might indirectly act as a strong selective force. These observations suggest an important role for the average temperature and precipitation in shaping genetic differentiation between locally adapted populations.

4.5.5 Evidence for selection on temperature-related genes

We analyzed the function of candidate loci under selection in response to climate variation in order to find cues regarding their phenotypic effect, and to provide insight into adaptive variation. In line with annual mean temperature explaining large part of adaptive variation, we find that the function of many candidate genes under selection are related to energy homeostasis and fat metabolism. *M. glareolus* populations across the European continent experience different ambient temperatures, which obviously results in varying energetic requirements throughout the species range. Maintaining a constant body temperature can be especially challenging in cold conditions, as these results in high thermogenic demands. *M. glareolus* are in general smaller towards higher latitudes and have a lower basal metabolic rate (BMR) (Ledevin et al. 2010; Boratyński et al. 2011). At the start of winter, individuals reduce their body mass and increase their capacity for non-shivering thermogenesis (NST) (Klaus et al.

1988). Which is presumed to play an important role in adaptation to colder conditions, as a lower body mass decreases the total cost of maintenance and thermoregulation (Klaus et al. 1988; Ergon et al. 2004; Boratyński et al. 2011; Boratyński et al. 2013). Adipose tissue plays an important role in energy homeostasis, regulates energy storage and expenditure and is a major determinant of body mass (Sethi & Vidal-Puig 2007). Of interest therefore are a number of genes involved in adipose biology and body size that were found to be under divergent selection. Examples of such genes are *leucine rich repeat containing 8 family, member C* (*Lrrc8c*) (Tominaga et al. 2004), *dynein axonemal heavy chain 8* (*Dnah8*) (Söhle et al. 2012) and *potassium voltage-gated channel subfamily H member 1* (*Kcnh1*) (Y.Y. Zhang et al. 2014), which is under selection in human Alaskan Arctic populations adapted to cold environments (Reynolds et al. 2019). *Insulin-like growth factor 1* (*Igf1*), encodes for a growth factor protein that is related to insulin but has a much higher growth-promoting activity. *Igf1* is strongly associated with body size in mammals such as mice (Kroonsberg et al. 1989; Baker et al. 1993), humans (Woods et al. 1996) and dogs (Sutter et al. 2007), possibly via its regulatory effects on cell growth and cellular differentiation. Furthermore, we found a signal of divergent selection in *neurotrophic receptor tyrosine kinase 2* (*Ntrk2*), which has been implicated in the regulation of food intake and body weight in rodents. (Kernie 2000; Xu et al. 2003). In general, genetic variation in these genes may be related to variation in body size and associated differences in adipose metabolism across *M. glareolus* distribution.

In mammals, heat can be produced via NST in the brown adipose which supports thermoregulation. NST can change in response to environmental conditions and is increased in winter (Klaus et al. 1988; Bonda-Ostaszewska et al. 2012). The capacity for NST differs between evolutionary lineages in *M. arvalis* (Bize et al. 2018), suggesting this trait is under selection, presumably as an adaptation to cold conditions. It is therefore interesting that we found a signal of selection in the *phospholipase C-like 1* (*Plcl1*) gene. The product of this gene is involved in several intracellular signaling cascades involved in adipose metabolism and is a modulator of lipolysis (Oue et al. 2016). In addition to this, it serves as a signaling molecule regulating NST (Kanematsu et al. 2019). Interestingly, three of these described genes (*Plcl1*, *Lrrc8c* and *Kcnh1*) were also implied to be involved in adaptation to climate in populations of *M. musculus* (Phifer-Rixey et al. 2018). Several of the above genes related to energy homeostasis are also associated with obesity in humans (*Lrrc8c* (Hayashi et al. 2011), *Plcl1* (Yamawaki et al. 2017), *Dnah8* (Söhle et al. 2012), *Kcnh1* (Vasconcelos et al. 2016), *Igf1*

(Berryman et al. 2013) and *Ntrk2* (Gray et al. 2007)), which provides further support for their role in maintaining energy homeostasis. Temperature is one of the most important environmental factors, affecting metabolism and the physiological of all organisms (Parmesan & Yohe 2003; Clarke 2003; Tewksbury et al. 2008). Numerous studies have used clinal variation in temperature and genome scans and detected signals of selection in genes related to energy homeostasis and metabolism (e.g. Andrew, Jensen, Hagen, Lundregan, & Griffith, 2018; Fumagalli et al., 2015; Hancock et al., 2011; S. E. Harris & Munshi-South, 2017; Harrisson et al., 2017; Lv et al., 2014; Prates, Penna, Rodrigues, & Carnaval, 2018; Pritchard et al., 2018; Rodríguez et al., 2017). This suggests that temperature is one of the key environmental variables driving local adaptation. It is therefore no surprise that we have also found genes related to adipose metabolism and energy homeostasis under divergent selection in response to variation in climate. The overlap between our candidate genes with candidate genes detected in *M. musculus* (Phifer-Rixey et al. 2018), suggests that small mammals might have a similar genetic basis for adapting to variation in temperature.

4.5.6 Evidence for selection on immune-related genes

Pathogens are among one of the strongest selective forces in natural populations and an important driver of local adaptation (Altizer et al. 2003; Fumagalli et al. 2011). There is large variation in the strength of pathogen-driven selection across geographic areas, caused by local variation in features such as climate, infection risk and resource availability (Charbonnel & Cosson 2012; McIntyre et al. 2017). Thus, given the wide distribution of sampled *M. glareolus* populations, the distribution of pathogens is likely highly heterogeneous among them (Tersago et al. 2009; Guivier et al. 2014). Previous studies have found evidence for adaptive variation associated with pathogen tolerance on a much smaller scale than our study (Guivier et al. 2010; Guivier et al. 2011; White et al. 2013; Guivier et al. 2014; Dubois et al. 2017; Rohfritsch et al. 2018). Accordingly, we found signals of divergent selection in a number of genes related to pathogen handling and immune functioning. Such as, *signal transducer and activator of transcription 4* (*Stat4*), which encodes a transcription factor responsible for T-helper cell development. *Stat4* plays a major role in the immune system response to viral infections as being part of the JAK-STAT signaling pathway (Kaplan 2005; Villarino et al. 2017). Interestingly, *Stat4* was found to be involved in the immune response to Sin Nombre hantavirus infected deer mice (*P. maniculatus*) (Schountz et al. 2012; Schountz et al. 2014). This suggests

that changes in this gene might play a role in PUUV resistance in *M. glareolus* as well. In addition to this, we found that two genes encoding transcription factor involved in pathogen handling were also under divergent selection. First, *Basic leucine zipper transcriptional factor ATF-like 3* (*BATF3*), is responsible for the cell fate choice of regulatory T-cells and antigen cross-presentation (Hildner et al. 2008; Geissmann et al. 2010; Lee et al. 2017). Second, we found a signal of selection in *B lymphoid transcription repressor BTB and CNC homology 2* (*Bach2*). Studies using mice have shown that *Bach2* is a key regulator of immune activation that prevents inflammatory disease by controlling the balance between tolerance and immunity (Roychoudhuri et al. 2013; Nakamura et al. 2013; Sidwell & Kallies 2016). Thus, it is possible that variation in this gene might play a role in balancing the inflammatory response towards PUUV in *M. glareolus* as well. The signals of divergent selection found in these immunology genes provide further evidence for the general assumption that pathogens play a major role as important drivers of local adaptation. Our findings are in line with previous results characterizing the genetic basis of *M. glareolus* pathogen tolerance abilities. Results suggest that divergent pathogen selection pressures across *M. glareolus* distribution has resulted in adaptive variation in contrasting environments.

4.6 Conclusion

In this study, we used ddRAD-sequencing and a combination of landscape genomic approaches to provide a view on the evolutionary processes acting across the *M. glareolus* distribution. We shed light on adaptive genetic variation by using univariate GEA-methods to detect outlier loci that correlate with climate. We expanded these methods with a multivariate approach that is able to pick up subtle signals of polygenic selection. After selecting outliers found by multiple methods, we identified loci in 74 genes of interest that showed evidence of spatially varying selection. Functional annotation of these suggest that spatially varying selection associated with energy homeostasis and responses to pathogens are among the more important ones throughout the *M. glareolus* distribution. In addition to this, we showed that both geography (measured by population structure) and climate play a large (and shared) role in explaining neutral genetic differentiation across the species range. Genetic variation among candidate loci was mostly explained by annual variation in climate and specifically annual mean temperature. Which highlights the importance of temperature as a key environmental factor driving adaptive variation in *M. glareolus*. The functional significance of the candidate genes found should be further investigates, as these loci are good candidates for adaptation to changing environmental

conditions. This could be done in future studies with additional populations and denser genotyping, or with experimental manipulation such as common-garden experiments combined with an analysis of gene expression. Understanding the balance between the strength of environmental selection and neutral processes can inform us about the spatial distribution of genetic variation and the genetic basis of local adaptation.

Chapter 5: Discussion

The study of local adaptation is central to the field of evolutionary biology. The advent of next generation sequencing has made it possible to obtain large-scale, genome-wide data of so far little studied wild species and populations. Voles are small mammals with a widespread distribution throughout Europe, where they encounter a wide variety of different environmental conditions. They, furthermore, have a high evolutionary potential which makes it likely that they genetically adapt to local conditions. These characteristics make them an ideal species to study the genetics of local adaptation. Although loci under natural selection have been identified in wild vole populations (Fischer et al. 2011; White et al. 2013; Fischer et al. 2014; Rohfritsch et al. 2018), our understanding of the genetic underpinnings of local adaptation in these species is still limited. Therefore, I set out to provide insight into the genetic basis of local adaptation in response to strong environmental heterogeneity. In this thesis, I provide evidence for the genetic underpinnings of adaptation to high altitude in *Microtus arvalis* (Chapter 3). In Chapter 4, I then identify adaptive loci associated with climate gradients in *Myodes glareolus* and highlight the importance of temperature in shaping adaptive variation. In addition, in Chapter 2, I present the first complete mitochondrial genome of *M. arvalis* and use this to decipher the phylogeny among this and related species.

5.1 Methodological considerations

5.1.1 ddRAD-sequencing

Reduced sequencing costs have enabled sequencing of entire populations of some organisms with extensive resources (Mackay et al. 2012; G. Zhang et al. 2014), however, sequencing whole genomes of non-model organisms still requires a large investment. I, therefore, chose to use ddRAD-sequencing (Peterson et al. 2012) as a cost-effective method to sequence parts of the genomes of the many individuals in chapter 3 and 4. However, several questions have been raised in regards to the functional significance of adaptive loci detected using reduced-representation methods that warrant discussion (Lowry et al. 2017). First, during the final PCR amplification step of the ddRAD-protocol, PCR duplicates, which are identical (non-independent) copies of the same molecule, are created. Some studies report that up to 30% of sequenced data in ddRAD libraries might be PCR duplicates (Schweyen et al. 2014; Andrews et al. 2014). Ideally, these would be identified by the fact that they map to the same start and end position of the genome, and removed in subsequent bioinformatic analyses as it is done in

genome shotgun studies. Unfortunately, this approach cannot be used with data from ddRAD sequencing, because all fragments from a given locus will have identical start and end positions as they are cut using the same enzymes. However, considering that fragments of a locus will have approximately the same length, it is unlikely that a substantial bias towards certain alleles exist, as there is no reason to assume PCR should favor one allele over another (Andrews et al. 2016). Nevertheless, as a conservative approach, to not over-amplify sequencing libraries and thereby create an excessive number of PCR duplicates, the optimal cycle numbers were determined based on qPCR amplification plots, which restricts the number of created PCR duplicates. Thus, although PCR duplicates cannot be avoided, it is unlikely that they are of major influence on the results of the studies in this thesis. Another concern has been raised about the ability of ddRAD to sequences all loci under selection across the genome (Lowry et al. 2017). The authors of this study argue that genome scans using RADseq will, depending on the level of linkage disequilibrium (LD), likely miss many markers under selection because only a fraction of the genome is sequenced using these approaches. For the two chapters in this thesis I generated data covering about 0.62% (chapter 3) and 0.33% (chapter 4) of the genome, resulting in one RAD-tag with a SNP approximately every 0.31 Mb and 0.25 Mb, respectively. However, the effectively covered region of the genome is much larger, as the signal from all loci in LD with our tags are also investigated. The expected coverage for RADseq data increases rapidly with LD (McKinney et al. 2017). For example, an average LD length of 250 kb would allow us to cover the whole genome of *M. arvalis* and *M. glareolus* (McKinney et al. 2017). Unfortunately, the extend of LD in both species is not known. In wild mice (*M. musculus domesticus*), which might have similar genetic characteristics as the studied species, LD decays at 100 kb (Laurie et al. 2007) Considering the large number of RAD-tags sequenced in both studies and the extend of LD in *M. musculus domesticus*, it can thus be assumed that the covered region of the genome is much larger than the region of the sequenced RAD-tags alone. Regardless of the actual genome coverage and although likely some loci under selection were missed, results of both studies show, that a large number of important loci under spatially varying natural selection were detected, with functions relevant to the selection pressures studied.

5.1.2 Genome scan methods

In chapter 3 and chapter 4 several different genome scan methods were used to identify loci potentially involved in local adaptation. These methods are widely used and have successfully

identified genes involved in local adaptation across a range of species (Rellstab et al. 2015; Hoban et al. 2016). However, they also all have their limitations. For example, a strong correlation between environmental gradients and demography history can lead to a lower power to detect loci under selection (de Villemereuil et al. 2014; Lotterhos & Whitlock 2015). In scenarios where theoretical assumptions regarding demographic history are not met, which often is the case in natural populations, this results in increased false discovery rates (de Villemereuil et al. 2014). The populations studied in both chapters are highly structured, and at least in the study presented in chapter 4, this structure correlates with the tested climate parameters. To account for this problem, I only used methods that correct for the confounding effects of population structure. The results obtained show that different methods detected different sets of outlier loci, which is not unexpected as it was previously reported in other studies (Lotterhos & Whitlock 2015). To increase the reliability of the detected loci, I therefore only considered loci as candidate loci if they were detected by at least two methods. As true positives are more likely to overlap between methods than false positives (de Villemereuil et al. 2014), this approach increases the likelihood that the candidate loci detected are indeed involved in local adaptation (François et al. 2016).

5.2 The first *Microtus arvalis* mitochondrial genome

M. arvalis is probably the most abundant European mammal and widely studied to understand postglacial recolonization processes of European fauna (Fink et al. 2004). However, despite its popularity as a research species, genomic resources for *M. arvalis* are mostly lacking. Mitochondrial DNA (mtDNA) is a popular and frequently used resource in evolutionary biology. It has been used extensively used as a marker of choice to resolve phylogenetic relationships and has provided tremendous insight into the process of evolutionary divergence (Avise et al. 1987; Galtier et al. 2009). However, until now, mitochondrial resources for *M. arvalis* were restricted to short, single gene or single region mitochondrial sequences. Those were used to determine the presence of several evolutionary lineages in *M. arvalis* (e.g. Fink et al., 2004; Heckel, Burri, Fink, Desmet, & Excoffier, 2005), but they were insufficient to fully resolve the relationship within the genus *Microtus* (Jaarola et al. 2004; Galewski et al. 2006). In the study presented in Chapter 2, shotgun sequencing data and an iterative mapping approach (Hahn et al. 2013) were used to recover the first complete mitochondrial genome of *M. arvalis* (Folkertsma et al. 2018). I used this to perform a phylogenetic analysis and to provide new insights into the Arvicolinae phylogeny. I did not find support for the previous generic

recognition of *Neodon* and *Lasiopodomys* as separate genera, suggesting these two genera have to be subsumed within the *Microtus* genus. The new mitochondrial sequence reported in this study provides an important genomic resource in this ecological relevant, but resource-wise under-represented species and can help to resolve uncertainty regarding the relationships between genera. This is illustrated by a recent publication (Takahata et al. 2019) that used the new *M. arvalis* mitochondrial genome to investigate the taxonomic status of a previously unknown species. Thus, the genomic resources generated within this study represent an important step forward as they can help to improve future phylogenetic analyses.

5.3 High-altitude adaptation in *Microtus arvalis*

In the second study, I set out to provide insights into the genetic basis of adaptation to high altitude. Organisms living at high altitude face considerable physiological challenges as a result of low temperatures, low oxygen levels and increased UV-radiation. These challenges are even more pronounced for small mammals, which, due to their small body size, are constantly confronted with high thermogenic demands. These high thermogenic demands require an adequate supply of oxygen to the tissue, which is challenging in the hypoxic conditions at high altitude. So far, our understanding about the genetic basis of high-altitude adaptation in mammals is mainly restricted to humans and deer mice (Bigham 2016; McClelland & Scott 2018; Storz et al. 2019), although other species have also been investigated to a limited extent (Scott et al. 2010; Natarajan et al. 2015; Waterhouse et al. 2018). Previous studies have provided evidence for adaptation to high altitude in *M. arvalis*, but the identification of genes with functional significance was hindered by the fact that only anonymous AFLP-markers were used (Fischer et al. 2011; Fischer et al. 2014). Hence, in Chapter 3, I aimed to provide insights into the genetic basis of adaptation to high altitude in *M. arvalis* to further our understanding of high-altitude adaptation. To this end, I analyzed genomic data generated using a ddRAD sequencing approach from individuals sampled at three low-altitude and three high-altitude study sites. I detected high levels of genetic differentiation among study sites, with a strong signal of isolation by distance (IBD). This resulted in a clear pattern of neutral population structure best represented by three major clusters, comprising of individuals from study sites in close proximity. These results are in line with previous studies in *M. arvalis* and are most likely a result of recurring population crashes and low effective gene flow between distant populations (Schweizer et al. 2007; Fischer et al. 2014). However, despite these strong patterns of neutral population structure, genome scan methods successfully detected signals of selection associated

with differences in altitude. Genome scan methods identified a number of candidate outlier genes with functions related to the formation of new blood vessels, the production of red blood cells and energy metabolism. This suggests that selection on high-altitude populations of *M. arvalis* is acting to increase oxygen delivery to the tissue and enhance thermogenic capacity. Moreover, a number of genes displaying signals of selection are directly involved or can be linked to the HIF-pathway. This ancient pathway is the major pathway that responds to changes in available oxygen in the cellular environment and regulates hundreds of downstream genes (Semenza 2007; Bigham & Lee 2014). It is widely implicated in high-altitude adaptation, and genes in this pathway have been under selection in several high-altitude species (e.g. humans, sheep, goats, dogs) (Simonson et al. 2010; Gou et al. 2014; Ai et al. 2014; Wei et al. 2016; Song et al. 2016). In addition, signals of selection were detected in genes that have been previously found to be under selection in high-altitude populations of other species. These results suggest that high-altitude populations of several taxa have evolved in a similar manner as a response to the unique conditions at high altitude. Thus, although the physiology of high-altitude adaptation is complex, it seems that natural selection has acted on only a small set of similar genes during the process of adaptation to high-altitude environments in different species.

The work presented in this chapter shows that even for non-model species without an annotated reference genome available, methods like ddRAD, can successfully be used to identify candidate genes for local adaptation. Researchers who generate RAD-sequencing data commonly resort to using *de novo* approaches when a reference genome for their study species is not available (Catchen et al. 2013; Puritz, Hollenbeck, et al. 2014; Eaton 2014). Such *de novo* approaches result in the generation of so-called *de novo* RAD assemblies. Those are reference genomes, consisting of several short contigs generated from RAD data. A huge drawback of this is that these *de novo* assemblies lack functional annotation. Thus, in the absence of genome annotation, a BLAST search has to be performed to annotate the generated *de novo* assembly. However, annotation results are often poor. As a consequence, the functional significance of a large percentage of loci of interest cannot be established, resulting in many loci of interest with unknown function as no annotation is available (ranging from 35%-95% of loci without annotation in studies (White et al. 2013; Guo et al. 2016; Rodríguez et al. 2017; Hoey & Pinsky 2018; Rohfritsch et al. 2018). Instead of using a *de novo* approach, reads were mapped to the well annotated prairie vole (*Microtus ochrogaster*) genome. Although *M. ochrogaster* diverged from *M. arvalis* about 13.2 MYA (Fabre et al. 2012), 83.4% of the reads were successfully

mapped to this reference genome. By using the positional information of outlier loci on the *M. ochrogaster* reference genome I was able to determine that 35% of outliers were found in annotated genes. Thus, our results demonstrate that a distant reference genome can be a great asset for identifying the functional significance of candidate genes in non-model organisms.

5.4 Climate adaptation in *Myodes glareolus*

Climate is of major influence on the distribution of species and affects all physiological processes of organisms (Parmesan & Yohe 2003; Clarke 2003). Detecting the genetic basis of adaptation to different climates and identifying climate related adaptive variation can provide insights into a species' ability to respond to climate change (Hoffmann & Willi 2008). Several studies have provided evidence for local adaptation associated with spatial variation in climate (Hereford 2009; Savolainen et al. 2013) and an increasing number of studies have focused on climate adaptation in non-model organisms (Haasl & Payseur 2016). There is also evidence for small mammals showing genetic changes related to variation in climate and other clinal gradients (e.g. Harris & Munshi-South, 2017; Phifer-Rixey et al., 2018; Rohfritsch et al., 2018; Waterhouse, Erb, Beever, & Russello, 2018). However, these studies often examined populations on relatively small geographic scales and our understanding about the genetic loci involved in climate adaptation in small mammals is thus rather limited. Moreover, little is known about the climatic forces that drive selective gradients and give rise to adaptive genetic variation for most species. Another drawback of most of studies published so far is that they have only used univariate methods to reveal adaptive variation. Therefore, they were rather limited in their ability to detect weak signals of polygenic selection (Wellenreuther & Hansson 2016). Hence, for the work presented in Chapter 4, I set out to provide insight into the genetic basis of climate adaptation in *M. glareolus*.

DdRAD sequencing was used to sequence and annotate the genome of 276 individuals from twelve *M. glareolus* populations across the European continent. Because of the wide distribution of sampled populations, climatic conditions varied widely among them. I first used several methods to characterize patterns of differentiation and population structure. These methods revealed high levels of differentiation among populations. Furthermore, a distinct pattern of population structure resembling the geographic distribution of populations was identified, in line with a strong signal of IBD. This was confirmed by results from the redundancy analysis (RDA), which demonstrate that the geographic distribution of populations

explains a large amount of genetic variation. RDA further uncovered a strong correlation between genetic population structure and geographic variation in climate, suggesting that a large amount of genetic variation was accounted for by spatially structured climate variation. However, in spite of the strong correlation between climate and population structure, the genome scan methods that were used still identified a large number of outlier loci that displayed signals of spatially varying selection associated with climate. Furthermore, the observation of outliers detected by the RDA suggests that at least some of the loci were under the influence of weak polygenic selection. I then used RDA to determine that annual mean temperature contributed most strongly to signatures of adaptive variation among outlier loci of all tested climatic variables, thus showing that temperature is an important selective pressure. This result is in accordance with previous research in *M. glareolus*, which showed seasonal and spatial phenotypic variation in temperature-related traits such as body mass, BMR and non-shivering thermogenesis (NST) (Klaus et al. 1988; Ledevin et al. 2010; Boratyński et al. 2011).

Signals of selection were found in a number of genes with functions related to energy homeostasis, body size, adipose metabolism and NST. The function of these genes suggests that these are differentially selected in response to the local environmental temperature of populations. This finding, in combination with the large amount of adaptive variation explained by temperature, indicates that temperature is an important selection pressure driving adaptation to local conditions in *M. glareolus* populations across their range. The observation that some of these genes are also implied to be involved in adaptation to climate in populations of *M. musculus* along a latitudinal gradient (Phifer-Rixey et al. 2018) suggests that small mammals might have a similar genetic basis for adapting to variation in temperature. I further identified various candidate genes with immunological functions, which suggests that populations experienced different pathogen selection pressures. Pathogens are among one of the strongest selective forces in natural populations and an important driver of local adaptation (Altizer et al. 2003; Fumagalli et al. 2011). The spatial distribution of pathogens is highly heterogeneous and many pathogens are climate sensitive (Guernier et al. 2004; McIntyre et al. 2017). *M. glareolus* is the reservoir host of the Puumala hantavirus (PUUV), but the prevalence of the virus differs among populations (Easterbrook & Klein 2008). Several studies have provided insight into the genetic basis of PUUV tolerance and found signals of selection in immune-related genes among populations that differ in PUUV prevalence on regional scales (Guivier et al. 2010; Guivier et al. 2014; Rohfritsch et al. 2018). In line with these studies, I identified signals of selection that

correlate with climate in immune-related genes. One of the candidate genes is involved in the response to Sin Nombre hantavirus in deer mice (Schountz et al. 2012; Schountz et al. 2014), while a further candidate gene is involved in regulating the balance between tolerance and immunity in the immune response (Roychoudhuri et al. 2013; Nakamura et al. 2013). These observations, along with the signals of selection identified in other immune-related genes, suggest that *M. glareolus* evolved specific genetic adaptations to pathogen selection pressures not only at a regional scale but across its entire distribution, possibly mediated by climate.

The results presented in this chapter show that improving the quality of genomic resources can facilitate downstream analysis and improve our understanding of adaptive variation in non-model organisms. The lack of a high-quality reference genome with proper genome annotation for *M. glareolus* initially hindered a thorough investigation of genes involved in climate adaptation. Although a reference genome was available (ASM130578v1), this consisted of many small scaffolds (below 1Kb in length) and no genome annotation was provided. To overcome this problem, I first set out to improve the contiguity of the reference genome. This was done by using the Cross-Species scaffolding pipeline (Grau et al. 2018). Here, I generated in silico mate-pair libraries using the reference genome from *M. ochrogaster* which diverged from *M. glareolus* about 14.2 MYA (Fabre et al. 2012). By doing so, the quality of the reference genome is greatly improved, which is reflected in the N50, which increased by a factor of 17. Using the improved reference genome resulted in a 14% increase of mapped reads compared to the original published *M. glareolus* reference genome. Then, as annotations were lacking, annotations for regions with outlier SNPs were obtained by aligning these regions to the well-annotated *M. ochrogaster* reference genome using the LastZ pairwise alignment tool (Harris, 2007). I was able to align more than 98% of the regions, which resulted in the successful identification of candidate genes. These results demonstrate that through the use of a specific set of software, I was able to improve genomic resources for this non-model species. This approach could be a promising avenue for studies on other species for which only little or only low-quality genomic resources are available, at least until high-quality genome assemblies are available for a much larger number of species than it is currently the case.

5.5 General conclusions and outlook

In this thesis I set out to provide novel insights into the genetic basis of local adaptation. I used ddRAD-sequencing, landscape genomic approaches and an innovative combination of methods to improve genomic resource, in order to identify candidate genes potentially involved in local adaptation. I studied vole populations at very different geographic scales. *M. arvalis* populations were studied across an altitudinal gradient in the Swiss Alps to identify adaptations to environmental conditions at high altitude. *M. glareolus* populations were studied across a much larger climate gradient in Europe to determine climate-related adaptations. Many of the genes identified in both studies have been found to be under selection in other species as a response to the same respective selection pressures. Not only does this increase confidence in our results, it also suggests that natural selection could act on similar sets of genes in the adaptive process of local adaptation in different species.

In many ways we have only begun to scratch the surface in our understanding of how species adapt to local conditions. We have identified adaptive loci putatively involved in local adaptation, but this is just the starting point of characterizing the genomic basis for local adaptation. With regard to the work presented here, future work should aim to verify the results presented in this thesis. The genetic basis of local adaptation often, but not always, varies among populations (Colosimo et al. 2005; Manceau et al. 2010; Turner et al. 2010), and studies which did use replicate tests determined that the overlap between replicates may be small (Poncet et al. 2010; Rellstab et al. 2017; Müller et al. 2017). If the loci identified in this thesis are truly adaptive, they should also show the same environmental correlations in independent datasets with different populations, which would improve evidence for the adaptive patterns discovered (Tiffin & Ross-Ibarra 2014; Rellstab et al. 2015). Moreover, high-quality genomic resources have to be developed for both species such as a well-annotated reference genome. This would provide a proper tool for future analysis, improving read alignment, variant calling and functional analysis, and would thereby aid our understanding of evolutionary processes in both species.

Studying the genetic basis of local adaptation does not end with identifying candidate loci. Research should continue from there and provide further evidence for the functional significance of candidate loci identified in genome scans. This can for instance be done by verifying if nonsynonymous SNPs are found in genes close to the identified loci (Hancock et

al. 2011) and if that is the case, by investigating their functional significance. Experimental validation with quantitative trait locus mapping of crosses in reciprocal transplant experiments can provide direct evidence that genetic differences in identified loci lead to a fitness advantage in affected local populations (Hancock et al. 2011; De Kort et al. 2014; Phifer-Rixey et al. 2018). This will improve our understanding of local adaptation, but the implementation of such approaches is not easily achieved and costs considerable effort.

Overall, the work presented within this thesis provides novel insights into the genetic basis of local adaptation in two species with little genomic resources, that have evolved adaptive changes in response to relevant environmental variation. With decreasing sequencing costs and an increase in the availability of genomic tools, promising avenues are opening up to improve our understanding of local adaptation in wild populations.

Bibliography

Aalto M, Górecki A, Meczeva R, Wallgren H, Weiner J. 1993. Metabolic rates of the bank voles (*Clethrionomys glareolus*) in Europe along a latitudinal gradient from Lapland to Bulgaria. Ann Zool Fennici. 30:233–238.

Aars J, Ims RA, Liu HP, Mulvey M, Smith MH. 1998. Bank voles in linear habitats show restricted gene flow as revealed by mitochondrial DNA (mtDNA). Mol Ecol. 7:1383–1389.

Abramson NI, Rodchenkova EN, Kostygov AY. 2009. Genetic variation and phylogeography of the bank vole (*Clethrionomys glareolus*, Arvicolinae, Rodentia) in Russia with special reference to the introgression of the mtDNA of a closely related species, red-backed vole (*Cl. rutilus*). Russ J Genet. 45:533–545.

Ai H, Yang B, Li J, Xie X, Chen H, Ren J. 2014. Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. BMC Genomics. 15:834.

Alexa J, Rahnenführer J. 2010. topGO: Enrichment analysis for gene ontology, R Package Version 2(0) (Bioconductor). :R package version 2.22.0.

Alkorta-Aranburu G, Beall CM, Witonsky DB, Gebremedhin A, Pritchard JK, Di Rienzo A. 2012. The genetic architecture of adaptations to high altitude in Ethiopia.Malik HS, editor. PLoS Genet. 8:e1003110.

Altizer S, Harvell D, Friedle E. 2003. Rapid evolutionary dynamics and disease threats to biodiversity. Trends Ecol Evol. 18:589–596.

Amirpour Haredasht S, Barrios M, Farifteh J, Maes P, Clement J, Verstraeten WW, Tersago K, van Ranst M, Coppin P, Berckmans D, Aerts JM. 2013. Ecological niche modelling of bank voles in Western Europe. Int J Environ Res Public Health. 10:499–514.

Anderson RA, Lovrien RE. 1984. Glycophorin is linked by band 4.1 protein to the human erythrocyte membrane skeleton. Nature. 307:655–658.

Andrew SC, Jensen H, Hagen IJ, Lundregan S, Griffith SC. 2018. Signatures of genetic adaptation to extremely varied Australian environments in introduced European house sparrows. Mol Ecol. 27:4542–4555.

Andrews KR, Good JM, Miller MR. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev. 17:81–92.

Andrews KR, Hohenlohe PA, Miller MR, Hand BK, Seeb JE, Luikart G. 2014. Trade-offs and utility of alternative RADseq methods: Reply to Puritz et al. Mol Ecol. 23:5943–6.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Mol Ecol. 22:3179–3190.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst. 18:489–522.

Babin C, Gagnaire PA, Pavey SA, Bernatchez L. 2017. RAD-Seq reveals patterns of additive polygenic variation caused by spatially-Varying selection in the American eel (*Anguilla rostrata*). Genome Biol Evol. 9:2974–2986.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 3:e3376.

Balzani E, Lassi G, Maggi S, Sethi S, Parsons MJ, Simon M, Nolan PM, Tucci V. 2016. The Zfhx3-Mediated Axis Regulates Sleep and Interval Timing in Mice. Cell Rep. 16:615–621.

Baker J, Liu JP, Robertson EJ, Efstratiadis A. 1993. Role of insulin-like growth factors in embryonic and postnatal growth. Cell. 75:73–82.

Baranano DE, Rao M, Ferris CD, Snyder SH. 2002. Biliverdin reductase: A major physiologic cytoprotectant. Proc Natl Acad Sci U S A. 99:16093–16098.

Barrett RDH, Schluter D. 2008. Adaptation from standing genetic variation. Trends Ecol Evol. 23:38–44.

Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Bely B, Bingley M, Bonilla C, Britto R, et al. 2017. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 45:D158–D169.

Baughman JM, Perocchi F, Girgis HS, Plovanich M, Belcher-Timme CA, Sancak Y, Bao XR, Strittmatter L, Goldberger O, Bogorad RL, et al. 2011. Integrative genomics identifies MCU as an essential component of the mitochondrial calcium uniporter. Nature. 476:341–345.

Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, Li C, Li JC, Liang Y, McCormack M, et al. 2010. Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders. Proc Natl Acad Sci U S A. 107:11459–11464.

Beall CM, Reichsman AB. 1984. Hemoglobin levels in a Himalayan high altitude population. Am J Phys Anthropol. 63:301–306.

Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol. 13:969–80.

Behnke JM, Barnard CJ, Bajer A, Bray D, Dinmore J, Frake K, Osmond J, Race T, Sinski E. 2001. Variation in the helminth community structure in bank voles (*Clethrionomys glareolus*) from three comparable localities in the mazury lake istrict region of Poland. Parasitology. 123:401–414.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Ser B. 57:289–300.

Bergmann C. 1848. Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse. [place unknown].

Bernardo R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci. 48:1649–1664.

Bernshtein AD, Apekina NS, Mikhailova T V., Myasnikov YA, Khlyap LA, Korotkov YS, Gavrilovskaya IN. 1999. Dynamics of Puumala hantavirus infection in naturally infected bank voles (*Clethrinomys glareolus*). Arch Virol. 144:2415–2428.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: Improved *de novo* metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 69:313–319.

Berryman DE, Glad CAM, List EO, Johannsson G. 2013. The GH/IGF-1 axis in obesity: Pathophysiology and therapeutic considerations. Nat Rev Endocrinol. 9:346–356.

Berthier K, Charbonnel N, Galan M, Chaval Y, Cosson JF. 2006. Migration and recovery of the genetic diversity during the increasing density phase in cyclic vole populations. Mol Ecol. 15:2665–2676.

Berven KA. 1982. The genetic basis of altitudinal variation in the wood frog *Rana sylvatica*. I. An experimental analysis of life history traits. Evolution (N Y). 36:962–983.

Beysard M, Heckel G. 2014. Structure and dynamics of hybrid zones at different stages of speciation in the common vole (*Microtus arvalis*). Mol Ecol. 23:673–687.

Bhagwat SA, Willis KJ. 2008. Species persistence in northerly glacial refugia of Europe: A matter of chance or biogeographical traits? J Biogeogr. 35:464–482.

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: The impact of rare variants. Genome Res. 23:1514–1521.

Bigham AW. 2016. Genetics of human origin and evolution: High-altitude adaptations. Curr Opin Genet Dev. 41:8–13.

Bigham AW, Lee FS. 2014. Human high-altitude adaptation: Forward genetics meets the HIF pathway. Genes Dev. 28:2189–2204.

Biyashev D, Veliceasa D, Kwiatek A, Sutanto MM, Cohen RN, Volpert O V. 2010. Natural angiogenesis inhibitor signals through Erk5 activation of peroxisome proliferator-activated receptor γ(PPARγ). J Biol Chem. 285:13517–13524.

Bize P, Lowe I, Lehto Hürlimann M, Heckel G. 2018. Effects of the mitochondrial and nuclear genomes on nonshivering thermogenesis in a wild derived rodent. Integr Comp Biol. 58:532–543.

Black CP, Tenney SM. 1980. Oxygen transport during progressive hypoxia in high-altitude and sea-level waterfowl. Respir Physiol. 39:217–239.

Blair LM, Granka JM, Feldman MW. 2014. On the stability of the Bayenv method in assessing human SNP-environment associations. Hum Genomics. 8:1.

Blanquart F, Kaltz O, Nuismer SL, Gandon S. 2013. A practical guide to measuring local adaptation. Ecol Lett. 16:1195–1205.

Blount ZD, Lenski RE, Losos JB. 2018. Contingency and determinism in evolution: Replaying life's tape. Science (80-). 362:eaam5979.

Bonda-Ostaszewska E, Włostowski T, Krasowska A, Kozłowski P. 2012. Seasonal and photoperiodic effects on lipid droplet size and lipid peroxidation in the brown adipose tissue of bank voles (*Myodes glareolus*). Acta Theriol (Warsz). 57:289–294.

Bondrup-Nielsen S, Ims RA. 1986. Comparison of maturation of female *Clethrionomys glareolus* from cyclic and noncyclic populations. Can J Zool. 64:2099–2102.

Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S, SanCristobal M. 2010. Detecting selection in population trees: The Lewontin and Krakauer test extended. Genetics. 186:241–262.

Bonin A. 2008. Population genomics: A new generation of genome scans to bridge the gap with functional genomics. Mol Ecol. 17:3583–3584.

Bonnet T, Leblois R, Rousset F, Crochet PA. 2017. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. Evolution (N Y). 71:2140–2158.

Boratyński Z, Alves PC, Berto S, Koskela E, Mappes T, Melo-Ferreira J. 2011. Introgression of mitochondrial DNA among *Myodes* voles: Consequences for energetics? BMC Evol Biol. 11:355.

Boratyński Z, Ketola T, Koskela E, Mappes T. 2016. The sex specific genetic variation of energetics in bank voles, consequences of introgression? Evol Biol. 43:37–47.

Boratyński Z, Koskela E, Mappes T, Schroderus E. 2013. Quantitative genetics and fitness effects of basal metabolism. Evol Ecol. 27:301–314.

Borkowska A, Ratkiewicz M, Jaarola M. 2010. Maintenance of genetic variation and moderate differentiation among populations under sex-biased dispersal in the common vole *Microtus arvalis* in patchy habitats. Acta Theriol (Warsz). 55:333–341.

Bourne EC, Bocedi G, Travis JMJ, Pakeman RJ, Brooker RW, Schiffers K. 2014. Between migration load and evolutionary rescue: Dispersal, adaptation and the response of spatially structured populations to environmental change. Proc R Soc B Biol Sci. 281:20132795.

Boyce CCK, Boyce JL. 1988. Population biology of Microtus arvalis. III. Regulation of numbers and breeding dispersion of females. J Anim Ecol. 57:737.

Braaker S, Heckel G. 2009. Transalpine colonisation and partial phylogeographic erosion by dispersal in the common vole (*Microtus arvalis*). Mol Ecol. 18:2518–2531.

Branco S, Bi K, Liao HL, Gladieux P, Badouin H, Ellison CE, Nguyen NH, Vilgalys R, Peay KG, Taylor JW, Bruns TD. 2017. Continental-level population differentiation and environmental adaptation in the mushroom *Suillus brevipes*. Mol Ecol. 26:2063–2076.

Bridle JR, Polechová J, Kawata M, Butlin RK. 2010. Why is adaptation prevented at ecological margins? New insights from individual-based simulations. Ecol Lett. 13:485–494.

Brommer JE, Pietiäinen H, Ahola K, Karell P, Karstinen T, Kolunen H. 2010. The return of the vole cycle in southern Finland refutes the generality of the loss of cycles through "climatic forcing." Glob Chang Biol. 16:577–586.

Buehler D, Holderegger R, Brodbeck S, Schnyder E, Gugerli F. 2014. Validation of outlier loci through replication in independent data sets: A test on *Arabis alpina*. Ecol Evol. 4:4296–4306.

Bullard RW, Broumand C, Meyer FR. 1966. Blood characteristics and volume in two rodents native to high altitude. J Appl Physiol. 21:994–998.

Bužan E V., Förster DW, Searle JB, Kryštufek B. 2010. A new cytochrome b phylogroup of the common vole (*Microtus arvalis*) endemic to the Balkans and its implications for the evolutionary history of the species. Biol J Linn Soc. 100:788–796.

Capblancq T, Luu K, Blum MGB, Bazin É. 2018. Evaluation of redundancy analysis to identify signatures of local adaptation. Mol Ecol Resour. 18:1223–1233.

Carleton MD, Gardner AL, Pavlinov IY, Musser GG. 2014. The valid generic name for red-backed voles (Muroidea: Cricetidae: Arvicolinae): Restatement of the case for *Myodes* Pallas, 1811. J Mammal. 95:943–959.

Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: An analysis tool set for population genomics. Mol Ecol. 22:3124–3140.

Cattell RB. 1966. The scree test for the number of factors. Multivariate Behav Res. 1:245–276.

Challinor AJ, Wheeler TR, Craufurd PQ, Ferro CAT, Stephenson DB. 2007. Adaptation of crops to climate change through genotypic responses to mean and extreme temperatures. Agric Ecosyst Environ. 119:190–204.

Charbonnel N, Cosson J-F. 2012. Molecular epidemiology of disease resistance genes with perspectives for researches on biological invasions and hybrid zones. In: Morand S, Beaudeau F, Cabaret J, editors. New Front Mol Epidemiol Infect Dis. Dordrecht: Springer Netherlands; p. 255–290.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics. 134:1289–1303.

Charlesworth B, Nordborg M, Charlesworth D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet Res. 70:155–174.

Chen LM, Kuo WW, Yang JJ, Wang SGP, Yeh YL, Tsai FJ, Ho YJ, Chang MH, Huang CY, Lee S Da. 2007. Eccentric cardiac hypertrophy was induced by long-term intermittent hypoxia in rats. Exp Physiol. 92:409–416.

Cheviron ZA, Bachman GC, Connaty AD, McClelland GB, Storz JF. 2012. Regulatory changes contribute to the adaptive enhancement of thermogenic capacity in high-altitude deer mice. Proc Natl Acad Sci U S A. 109:8635–8640.

Cheviron ZA, Brumfield RT. 2012. Genomic insights into adaptation to high-altitude environments. Heredity (Edinb). 108:354–361.

Cheviron ZA, Connaty AD, McClelland GB, Storz JF. 2014. Functional genomics of adaptation to hypoxic cold-stress in high-altitude deer mice: Transcriptomic plasticity and thermogenic performance. Evolution (N Y). 68:48–62.

Clarke A. 2003. Costs and consequences of evolutionary temperature adaptation. Trends Ecol Evol. 18:573–581.

Colangelo P, Aloise G, Franchini P, Annesi F, Amori G. 2012. Mitochondrial DNA reveals hidden diversity and an ancestral lineage of the bank vole in the Italian peninsula.Volff J-N, editor. J Zool. 287:41–52.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson H, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science (80-). 307:1928–1933.

Conover DO, Duffy TA, Hice LA. 2009. The covariance between genetic and environmental influences across ecological gradients. Ann N Y Acad Sci. 1168:100–129.

Cook LM, Saccheri IJ. 2013. The peppered moth and industrial melanism: Evolution of a natural selection case study. Heredity (Edinb). 110:207–212.

Coop G, Witonsky DB, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics. 185:1411–1423.

Cooper EA, Uy JAC. 2017. Genomic evidence for convergent evolution of a key trait underlying divergence in island birds. Mol Ecol. 26:3760–3774.

Corbet GB. 1964. Regional variation in the bank vole *Clethrionomys glareolus* in the British Isles. Proc R Soc B Biol Sci. 143:191–219.

Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. Mol Ecol. 21:1548–1566.

Cresko WA, Amores A, Wilson C, Murphy J, Currey M, Phillips P, Bell MA, Kimmel CB, Postlethwait JH. 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proc Natl Acad Sci U S A. 101:6050–6055.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol. 23:3133–3157.

Danecek P, Auton A, Abecasis GR, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–8.

David JR, Araripe LO, Chakir M, Legout H, Lemos B, Pétavy G, Rohmer C, Joly D, Moreteau B. 2005. Male sterility at extreme temperatures: A significant but neglected phenomenon for understanding *Drosophila* climatic adaptations. J Evol Biol. 18:838–846.

Deffontaine V, Libois R, Kotlík P, Sommer R, Nieberding C, Paradis E, Searle JB, Michaux JR. 2005. Beyond the Mediterranean peninsulas: Evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). Mol Ecol. 14:1727–1739.

DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43:491–501.

Deter J, Bryja J, Chaval Y, Galan M, Henttonen H, Laakkonen J, Voutilainen L, Vapalahti O, Vaheri A, Salvador AR, et al. 2008. Association between the DQA MHC class II gene and Puumala virus infection in *Myodes glareolus*, the bank vole. Infect Genet Evol. 8:450–458.

Devlin B, Roeder K. 1999. Genomic control for association studies. Biometrics. 55:997–1004.

Diamond SE, Chick L, Perez A, Strickler SA, Martin RA. 2017. Rapid evolution of ant thermal tolerance across an urban-rural temperature cline. Biol J Linn Soc. 121:248–257.

Diamond SE, Martin RA. 2016. The interplay between plasticity and evolution in response to human-induced environmental change. F1000Research. 5:2835.

Dimond JL, Gamblewood SK, Roberts SB. 2017. Genetic and epigenetic insight into morphospecies in a reef coral. Mol Ecol. 26:5031–5042.

Dixon P. 2003. VEGAN, a package of R functions for community ecology. J Veg Sci. 14:927–930.

Doebele RC, Schulze-Hoepfner FT, Hong J, Chlenski A, Zeitlin BD, Goel K, Gomes S, Liu Y, Abe MK, Nor JE, et al. 2009. A novel interplay between Epac/Rap1 and mitogen-activated protein kinase kinase 5/extracellular signal-regulated kinase 5 (MEK5/ERK5) regulates thrombospondin to control angiogenesis. Blood. 114:4592–4600.

Drewes S, Sheikh Ali H, Saxenhofer M, Rosenfeld UM, Binder F, Cuypers F, Schlegel M, Röhrs S, Heckel G, Ulrich RG. 2017. Host-associated absence of human Puumala virus infections in northern and Eastern Germany. Emerg Infect Dis. 23:83–86.

Dubois A, Galan M, Cosson JF, Gauffre B, Henttonen H, Niemimaa J, Razzauti M, Voutilainen L, Vitalis R, Guivier E, Charbonnel N. 2017. Microevolution of bank voles (*Myodes glareolus*) at neutral and immune-related genes during multiannual dynamic cycles: Consequences for Puumala hantavirus epidemiology. Infect Genet Evol. 49:318–329.

Easterbrook JD, Klein SL. 2008. Immunological mechanisms mediating hantavirus persistence in rodent reservoirs.Buchmeier MJ, editor. PLoS Pathog. 4:e1000172.

Eaton DAR. 2014. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics. 30:1844–1849.

Eccard JA, Herde A. 2013. Seasonal variation in the behaviour of a short-lived rodent. BMC Ecol. 13:1–9.

Eccard JA, Ylönen H. 2011. Initiation of breeding after winter in bank voles: Effects of food and population density. Can J Zool. 79:1743–1753.

Edmonds CA, Lillie AS, Cavalli-Sforza LL. 2004. Mutations arising in the wave front of an expanding population. Proc Natl Acad Sci U S A. 101:975–979.

Ehinger M, Fontanillas P, Petit E, Perrin N. 2002. Mitochondrial DNA variation along an altitudinal gradient in the greater white-toothed shrew, *Crocidura russula*. Mol Ecol. 11:939–945.

Ergon T, Speakman JR, Scantlebury M, Cavanagh R, Lambin X. 2004. Optimal body size and energy expenditure during winter: Why are voles smaller in declining populations? Am Nat. 163:442–457.

Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. Heredity (Edinb). 103:285–298.

Excoffier L, Ray N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol. 23:347–351.

Fabre P-H, Hautier L, Dimitrov D, Douzery E. 2012. A glimpse on the pattern of rodent diversification: A phylogenetic approach. BMC Evol Biol. 12:88.

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350.

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int Journal Climatol 37:4302–4315.

Filipi K, Marková S, Searle JB, Kotlík P. 2015. Mitogenomic phylogenetics of the bank vole *Clethrionomys glareolus*, a model system for studying end-glacial colonization of Europe. Mol Phylogenet Evol. 82, Part A:245–257.

Fink S, Excoffier L, Heckel G. 2004. Mitochondrial gene diversity in the common vole *Microtus arvalis* shaped by historical divergence and local adaptations. Mol Ecol. 13:3501–3514.

Fink S, Fischer MC, Excoffier L, Heckel G. 2010. Genomic scans support repetitive continental colonization events during the rapid radiation of voles (Rodentia: *Microtus*): the utility of AFLPs versus mitochondrial and nuclear sequence markers. Syst Biol. 59:548–572.

Fischer MC, Foll M, Excoffier L, Heckel G. 2011. Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). Mol Ecol. 20:1450–1462.

Fischer MC, Foll M, Heckel G, Excoffier L. 2014. Continental-scale footprint of balancing and positive selection in a small rodent (*Microtus arvalis*). PLoS One. 9:e112332.

Folkertsma R, Westbury M V., Eccard JA, Hofreiter M. 2018. The complete mitochondrial genome of the common vole, *Microtus arvalis* (Rodentia: Arvicolinae). Mitochondrial DNA Part B. 3:446–447.

Foll M, Gaggiotti OE. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics. 180:977–993.

Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L. 2014. Widespread signals of convergent adaptation to high altitude in Asia and America. Am J Hum Genet. 95:394–407.

Fontanillas P, Dépraz A, Giorgi MS, Perrin N. 2005. Nonshivering thermogenesis capacity associated to mitochondrial DNA haplotypes and gender in the greater white-toothed shrew, *Crocidura russula*. Mol Ecol. 14:661–670.

Forester BR, Lasky JR, Wagner HH, Urban DL. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Mol Ecol. 27:2215–2233.

François O, Martins H, Caye K, Schoville SD. 2016. Controlling false discoveries in genome scans for selection. Mol Ecol. 25:454–469.

Frankham R. 1995. Inbreeding and extinction: A threshold effect. Conserv Biol. 9:792–799.

Franks SJ, Hoffmann AA. 2012. Genetics of climate change adaptation. Genetics. 46.

Frichot E, François O. 2015. LEA: An R package for landscape and ecological association studies. Methods Ecol Evol.:n/a-n/a.

Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. 2014. Fast and efficient estimation of individual ancestry coefficients. Genetics. 196:973–983.

Frichot E, Schoville SD, Bouchard G, François O. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol. 30:1687–1699.

Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, Korneliussen TS, Gerbault P, Skotte L, Linneberg A, et al. 2015. Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science (80-). 349:1343–1347.

Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admettla A, Pattini L, Nielsen R. 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution.Akey JM, editor. PLoS Genet. 7:e1002355.

Fumagalli M, Vieira FG, Linderoth T, Nielsen R. 2014. NgsTools: Methods for population genetics analyses from next-generation sequencing data. Bioinformatics. 30:1486–1487.

Galewski T, Tilak MK, Sanchez S, Chevret P, Paradis E, Douzery E. 2006. The evolutionary radiation of Arvicolinae rodents (voles and lemmings): Relative contribution of nuclear and mitochondrial DNA phylogenies. BMC Evol Biol. 6:80.

Le Galliard J-F, Rémy A, Ims RA, Lambin X. 2012. Patterns and processes of dispersal behaviour in arvicoline rodents. Mol Ecol. 21:505–23.

Galtier N, Nabholz B, Glémin S, Hurst GDD. 2009. Mitochondrial DNA as a marker of molecular diversity: A reappraisal. Mol Ecol. 18:4541–4550.

Gandon S, Michalakis Y. 2002. Local adaptation, evolutionary potential and host-parasite coevolution: Interactions between migration, mutation, population size and generation time. J Evol Biol. 15:451–462.

Gansauge MT, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat Protoc. 8:737–748.

Garant D, Forde SE, Hendry AP. 2007. The multifarious effects of dispersal and gene flow on contemporary adaptation. Funct Ecol. 21:434–443.

García JT, Domínguez-Villaseñor J, Alda F, Calero-Riestra M, Pérez Olea P, Fargallo JA, Martínez-Padilla J, Herranz J, Oñate JJ, Santamaría A, et al. 2019. A complex scenario of glacial survival in Mediterranean and continental refugia of a temperate continental vole species (*Microtus arvalis*) in Europe. J Zool Syst Evol Res.:jzs.12323.

Gauffre B, Berthier K, Inchausti P, Chaval Y, Bretagnolle V, Cosson JF. 2014. Short-term variations in gene flow related to cyclic density fluctuations in the common vole. Mol Ecol. 23:3214–3225.

Gauffre B, Petit E, Brodier S, Bretagnolle V, Cosson JF. 2009. Sex-biased dispersal patterns depend on the spatial scale in a social rodent. Proc R Soc B Biol Sci. 276:3487–3494.

Gautier M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics. 201:1555–1579.

Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, Cornuet JM, Estoup A. 2013. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Mol Ecol. 22:3165–3178.

Geissmann F, Manz MG, Jung S, Sieweke MH, Merad M, Ley K. 2010. Development of monocytes, macrophages, and dendritic cells. Science (80-). 327:656–661.

Geraldes A, Farzaneh N, Grassa CJ, Mckown AD, Guy RD, Mansfield SD, Douglas CJ, Cronk QCB. 2014. Landscape genomics of *populus trichocarpa*: The role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. Evolution (N Y). 68:3260–3280.

Gerlach G, Musolf K. 2000. Fragmentation of landscape as a cause for genetic subdivision in bank voles. Conserv Biol. 14:1066–1074.

Getz LL, Oli MK, Hofmann JE, McGuire B. 2007. Vole population dynamics: Factors affecting peak densities and amplitudes of annual population fluctuations of Microtus pennsylvanicus. Acta Theriol (Warsz). 52:159–170.

Gherardi G, Nogara L, Ciciliot S, Fadini GP, Blaauw B, Braghetta P, Bonaldo P, De Stefani D, Rizzuto R, Mammucari C. 2019. Loss of mitochondrial calcium uniporter rewires skeletal muscle metabolism and substrate preference. Cell Death Differ. 26:362–381.

Ghorbel MT, Cherif M, Jenkins E, Mokhtari A, Kenny D, Angelini GD, Caputo M. 2010. Transcriptomic analysis of patients with tetralogy of Fallot reveals the effect of chronic hypoxia on myocardial gene expression. J Thorac Cardiovasc Surg. 140:337-345.e26.

Ghorbel MT, Coulson JM, Murphy D. 2003. Cross-talk between hypoxic and circadian pathways: Cooperative roles for hypoxia-inducible factor 1α and CLOCK in transcriptional activation of the vasopressin gene. Mol Cell Neurosci. 22:396–404.

Gienapp P, Teplitsky C, Alho JS, Mills JA, Merilä J. 2008. Climate change and evolution: disentangling environmental and genetic responses. Mol Ecol. 17:167–178.

Gou X, Wang Z, Li N, Qiu F, Xu Z, Yan D, Yang S, Jia, Konga X, Wei Z, et al. 2014. Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. Genome Res. 24:1308–1315.

Graham CF, Glenn TC, McArthur AG, Boreham DR, Kieran T, Lance S, Manzon RG, Martino JA, Pierson T, Rogers SM, et al. 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Mol Ecol Resour. 15:1304–1315.

Grau JH, Hackl T, Koepfli KP, Hofreiter M. 2018. Improving draft genome contiguity with reference-derived in silico mate-pair libraries. Gigascience. 7.

Gray J, Yeo G, Hung C, Keogh J, Clayton P, Banerjee K, McAulay A, O'Rahilly S, Farooqi IS. 2007. Functional characterization of human NTRK2 mutations identified in patients with severe early-onset obesity. Int J Obes. 31:359–364.

Grundy GJ, Rulten SL, Zeng Z, Arribas-Bosacoma R, Iles N, Manley K, Oliver A, Caldecott KW. 2013. APLF promotes the assembly and activity of non-homologous end joining protein complexes. EMBO J. 32:112–125.

Guernier V, Hochberg ME, Guégan JF. 2004. Ecology drives the worldwide distribution of human diseases.Paul Harvey, editor. PLoS Biol. 2:e141.

Guivier E, Galan M, Chaval Y, Xuéreb A, Ribas Salvador A, Poulle M, Voutilainen L, Henttonen H, Charbonnel N, Cosson JF. 2011. Landscape genetics highlights the role of bank vole metapopulation dynamics in the epidemiology of Puumala hantavirus. Mol Ecol. 20:3569–3583.

Guivier E, Galan M, Henttonen H, Cosson JF, Charbonnel N. 2014. Landscape features and helminth co-infection shape bank vole immunoheterogeneity, with consequences for Puumala virus epidemiology. Heredity (Edinb). 112:274–281.

Guivier E, Galan M, Salvador AR, Xuéreb A, Chaval Y, Olsson GE, Essbauer SS, Henttonen H, Voutilainen L, Cosson JF, Charbonnel N. 2010. Tnf-α expression and promoter sequences reflect the balance of tolerance/resistance to Puumala hantavirus infection in European bank vole populations. Infect Genet Evol. 10:1208–1217.

Günther T, Coop G. 2013. Robust identification of local adaptation from allele frequencies. Genetics. 195:205–220.

Guo B, DeFaveri J, Sotelo G, Nair A, Merilä J. 2015. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. BMC Biol. 13:19.

Guo B, Lu D, Liao WB, Merilä J. 2016. Genomewide scan for adaptive differentiation along altitudinal gradient in the Andrew's toad Bufo andrewsi. Mol Ecol. 25:3884–3900.

Haasl RJ, Payseur BA. 2016. Fifteen years of genomewide scans for selection: Trends, lessons and unaddressed genetic sources of complication. Mol Ecol. 25:5–23.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - A baiting and iterative mapping approach. Nucleic Acids Res. 41:e129–e129.

Hahne J, Jenkins T, Halle S, Heckel G. 2011. Establishment success and resulting fitness consequences for vole dispersers. Oikos. 120:95–105.

Haldane JBS. 1948. The theory of a cline. J Genet. 48:277–284.

Hamilton G, Currat M, Ray N, Heckel G, Beaumont MA, Excoffier L. 2005. Bayesian estimation of recent migration rates after a spatial expansion. Genetics. 170:409–417.

Han E, Sinsheimer JS, Novembre J. 2014. Characterizing bias in population genetic inferences from low-coverage sequencing data. Mol Biol Evol. 31:723–35.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011. Adaptation to climate across the Arabidopsis thaliana genome. Science (80-). 334:83–86.

Hansson L. 1985. Geographic differences in bank voles *Clethrionomys glareolus* in relation to ecogeographical rules and possible demographic and nutritive strategies. Ann Zool Fennici. 22:319–328.

Hansson L, Henttonen H. 1985. Regional differences in cyclicity and reproduction in *Clethrionomys* species: Are they related? Ann Zool Fennici.:277–288.

Harris RS. 2007. Improved pairwise alignment of genomic DNA Ph. D. [place unknown].

Harris SE, Munshi-South J. 2017. Signatures of positive selection and local adaptation to urbanization in white-footed mice (*Peromyscus leucopus*). Mol Ecol. 26:6336–6350.

Harrisson KA, Amish SJ, Pavlova A, Narum SR, Telonis-Scott M, Rourke ML, Lyon J, Tonkin Z, Gilligan DM, Ingram BA, et al. 2017. Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity. Mol Ecol. 26:6253–6269.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P. 2014. Using genomics to characterize evolutionary potential for conservation of wild populations. Evol Appl. 7:1008–1025.

Hayashi T, Nozaki Y, Nishizuka M, Ikawa M, Osada S, Imagawa M. 2011. Factor for adipocyte differentiation 158 gene disruption prevents the body weight gain and insulin resistance induced by a high-fat diet. Biol Pharm Bull. 34:1257–63.

Hayes JP. 1989. Field and maximal metabolic rates of deer mice (*Peromyscus maniculatus*) at low and high altitudes. Physiol Zool. 62:732–744.

Haynes S, Jaarola M, Searle JB. 2003. Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. Mol Ecol. 12:951–956.

Heckel G, Burri R, Fink S, Desmet J-F, Excoffier L. 2005. Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. Evolution (N Y). 59:2231–2242.

Hedrick PW. 2011. Genetics of populations. [place unknown]: Jones and Bartlett Publishers.

Hereford J. 2009. A quantitative survey of local adaptation and fitness trade-offs. Am Nat. 173:579–588.

Hewitt GM. 1999. Post-glacial re-colonization of European biota. Biol J Linn Soc. 68:87–112.

Hildner K, Edelson BT, Purtha WE, Diamond M, Matsushita H, Kohyama M, Calderon B, Schraml BU, Unanue ER, Diamond MS, et al. 2008. Batf3 deficiency reveals a critical role for CD8α + dendritic cells in cytotoxic T cell immunity. Science (80-). 322:1097–1100.

Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd GS, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC. 2016. Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. Am Nat. 188:379–397.

Hoey JA, Pinsky ML. 2018. Genomic signatures of environmental selection despite near-panmixia in summer flounder. Evol Appl. 11:1732–1747.

Hof AEV t., Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 534:102–105.

Hofer T, Ray N, Wegmann D, Excoffier L. 2009. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. Ann Hum Genet. 73:95–108.

Hoffmann AA, Anderson A, Hallas R. 2002. Opposing clines for high and low temperature resistance in *Drosophila melanogaster*. Ecol Lett. 5:614–618.

Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. Nature. 470:479–485.

Hoffmann AA, Willi Y. 2008. Detecting genetic responses to environmental change. Nat Rev Genet. 9:421–432.

Holliday JA, Yuen M, Ritland K, Aitken SN. 2010. Postglacial history of a widespread conifer produces inverse clines in selective neutrality tests. Mol Ecol. 19:3857–3864.

Homola JJ, Loftin CS, Cammen KM, Helbing CC, Birol I, Schultz TF, Kinnison MT. 2019. Replicated landscape genomics identifies evidence of local adaptation to urbanization in wood frogs. J Hered. esz041.

Hu Z-L, Bao J, Reecy JM. 2008. CateGOrizer: A web-based program to batch analyze gene ontology classification categories. Online J Bioinforma. 9:108–112.

Huang H, Knowles L. 2016. Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of rad sequences. Syst Biol. 65:357–365.

Huang F, Wang X, Ostertag EM, Nuwal T, Huang B, Jan YN, Basbaum AI, Jan LY. 2013. TMEM16C facilitates Na(+)-activated K+ currents in rat sensory neurons and regulates pain processing. Nat Neurosci. 16:1284–1290.

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics. 132:583–9.

Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L. 2000. Rapid evolution of a geographic cline in size in an introduced fly. Science (80-). 287:308–309.

Imholt C, Reil D, Eccard JA, Jacob D, Hempelmann N, Jacob J. 2015. Quantifying the past and future impact of climate on outbreak patterns of bank voles (*Myodes glareolus*). Pest Manag Sci. 71:166–172.

Inchausti P, Carslake D, Attié C, Bretagnolle V. 2009. Is there direct and delayed density dependent variation in population structure in a temperate European cyclic vole population? Oikos. 118:1201–1211.

Jaarola M, Martínková N, Gündüz İ, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedakis-Tsolis S, et al. 2004. Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. Mol Phylogenet Evol. 33:647–663.

Jeffreys H. 1961. The theory of probability. 3rd ed. Oxford: Clarendon Press.

Jensen LD, Cao Y. 2013. Clock controls angiogenesis. Cell Cycle. 12:405–408.

Jensen LD, Cao Z, Nakamura M, Yang Y, Bräutigam L, Andersson P, Zhang Y, Wahlberg E, Länne T, Hosaka K, Cao Y. 2012. Opposing effects of circadian clock genes Bmal1 and Period2 in regulation of VEGF-dependent angiogenesis in developing zebrafish. Cell Rep. 2:231–241.

Johnson PLF, Slatkin M. 2008. Accounting for bias from sequencing error in population genetic estimates. Mol Biol Evol. 25:199–206.

De Jong MA, Collins S, Beldade P, Brakefield PM, Zwaan BJ. 2013. Footprints of selection in wild populations of *Bicyclus anynana* along a latitudinal cline. Mol Ecol. 22:341–353.

Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G, Taberlet P. 2007. A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. Mol Ecol. 16:3955–3969.

Kallio ER, Helle H, Koskela E, Mappes T, Vapalahti O. 2015. Age-related effects of chronic hantavirus infection on female host fecundity.Childs D, editor. J Anim Ecol. 84:1264–1272.

Kamer KJ, Mootha VK. 2015. The molecular era of the mitochondrial calcium uniporter. Nat Rev Mol Cell Biol. 16:545–553.

Kane NC, King MG, Barker MS, Raduski A, Karrenberg S, Yatabe Y, Knapp SJ, Rieseberg LH. 2009. Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *helianthus* species. Evolution (N Y). 63:2061–2075.

Kanematsu T, Oue K, Okumura T, Harada K, Yamawaki Y, Asano S, Mizokami A, Irifune M, Hirata M. 2019. Phospholipase C-related catalytically inactive protein: A novel signaling molecule for modulating fat metabolism and energy expenditure. J Oral Biosci. 61:65–72.

Kaplan MH. 2005. STAT4: A critical regulator of inflammation *in vivo*. Immunol Res. 31:231–241.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 30:772–780.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. Ecol Lett. 7:1225–1241.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649.

Keller I, Alexander JM, Holderegger R, Edwards PJ. 2013. Widespread phenotypic and genetic divergence along altitudinal gradients in animals. J Evol Biol. 26:2527–2543.

Kernie SG. 2000. BDNF regulates eating behavior and locomotor activity in mice. EMBO J. 19:1290–1300.

Kim EJ, Yoo YG, Yang WK, Lim YS, Na TY, Lee IK, Lee MO. 2008. Transcriptional activation of HIF-1 by RORα and its role in hypoxia signaling. Arterioscler Thromb Vasc Biol. 28:1796–1802.

Kjeldsen SR, Zenger KR, Leigh K, Ellis W, Tobey J, Phalen D, Melzer A, FitzGibbon S, Raadsma HW. 2016. Genome-wide SNP loci reveal novel insights into koala (*Phascolarctos cinereus*) population variability across its range. Conserv Genet. 17:337–353.

Klaus S, Heldmaier G, Ricquier D. 1988. Seasonal acclimation of bank voles and wood mice: Nonshivering thermogenesis and thermogenic properties of brown adipose tissue mitochondria. J Comp Physiol B. 158:157–164.

Kloch A, Babik W, Bajer A, Siński E, Radwan J. 2010. Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*. Mol Ecol. 19:255–265.

Klopfstein S, Currat M, Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. Mol Biol Evol. 23:482–490.

Konczal M, Babik W, Radwan J, Sadowska ET, Koteja P. 2015. Initial molecular-level response to artificial selection for increased aerobic metabolism occurs primarily through changes in gene expression. Mol Biol Evol. 32:1461–1473.

Konczal M, Koteja P, Orlowska-Feuer P, Radwan J, Sadowska ET, Babik W. 2016. Genomic response to selection for predatory behavior in a mammalian model of adaptive radiation. Mol Biol Evol. 33:2429–2440.

Konczal M, Koteja P, Stuglik MT, Radwan J, Babik W. 2014. Accuracy of allele frequency estimation using pooled RNA-Seq. Mol Ecol Resour. 14:381–392.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 15:356.

Korneliussen TS, Moltke I. 2015. NgsRelate: A software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics. 31:4009–4011.

Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R. 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinformatics. 14:289.

Körner C. 2007. The use of 'altitude' in ecological research. Trends Ecol Evol. 22:569–574.

De Kort H, Vandepitte K, Bruun HH, Closset-Kopp D, Honnay O, Mergeay J. 2014. Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. Mol Ecol. 23:4709–4721.

Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: A review. Plant Methods. 9:29.

Kotlík P, Deffontaine V, Mascheretti S, Zima J, Michaux JR, Searle JB. 2006. A northern glacial refugium for bank voles (*Clethrionomys glareolus*). Proc Natl Acad Sci U S A. 103:14860–14864.

Kotlík P, Marková S, Konczal M, Babik W, Searle JB. 2018. Genomics of end-Pleistocene population replacement in a small mammal. Proc R Soc B Biol Sci. 285:20172624.

Kotlík P, Marková S, Vojtek L, Stratil A, Šlechta V, Hyršl P, Searle JB, Slechta V. 2014. Adaptive phylogeography: Functional divergence between haemoglobins derived from different glacial refugia in the bank vole. Proc R Soc B Biol Sci. 281:20140021.

Kroonsberg C, McCutcheon SN, Siddiqui RA, Mackenzie DD, Blair HT, Ormsby JE, Breier BH, Gluckman PD. 1989. Reproductive performance and fetal growth in female mice from lines divergently selected on the basis of plasma IGF-1 concentrations. J Reprod Fertil. 87:349–53.

Kwong JQ, Huo J, Bround MJ, Boyer JG, Schwanekamp JA, Ghazal N, Maxwell JT, Jang YC, Khuchua Z, Shi K, et al. 2018. The mitochondrial calcium uniporter underlies metabolic fuel preference in skeletal muscle. JCI Insight. 3.

Lambin X, Bretagnolle V, Yoccoz NG. 2006. Vole population cycles in northern and southern Europe: Is there a need for different explanations for single pattern? J Anim Ecol. 75:340–349.

Lande R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution (N Y). 30:314.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol. 29:1695–1701.

Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH. 2012. Characterizing genomic variation of *Arabidopsis thaliana*: The roles of geography and climate. Mol Ecol. 21:5512–5529.

Lau DS, Connaty AD, Mahalingam S, Wall N, Cheviron ZA, Storz JF, Scott GR, McClelland GB. 2017. Acclimation to hypoxia increases carbohydrate use during exercise in high-altitude deer mice. Am J Physiol - Regul Integr Comp Physiol. 312:R400–R411.

Laurie CC, Nickerson DA, Anderson AD, Weir BS, Livingston RJ, Dean MD, Smith KL, Schadt EE, Nachman MW. 2007. Linkage disequilibrium in wild mice. PLoS Genet. 3:e144.

Ledevin R, Chevret P, Helvaci Z, Michaux JR, Renaud S. 2018. Bank voles in Southern Eurasia: Vicariance and adaptation. J Mamm Evol. 25:119–129.

Ledevin R, Michaux JR, Deffontaine V, Henttonen H, Renaud S. 2010. Evolutionary history of the bank vole *Myodes glareolus*: A morphometric perspective. Biol J Linn Soc. 100:681–694.

Lee W, Kim HS, Hwang SS, Lee GR. 2017. The transcription factor batf3 inhibits the differentiation of regulatory T cells in the periphery. Exp Mol Med. 49:e393–e393.

Legendre P, Legendre LFJ. 2012. Numerical ecology. [place unknown]: Elsevier.

Lenormand T. 2002. Gene flow and the limits to natural selection. Trends Ecol Evol. 17:183–189.

Lepais O, Weir JT. 2014. SimRAD: An R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol Ecol Resour. 14:1314–1321.

Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T. 2009. Genome-wide association study of blood pressure and hypertension. Nat Genet. 41:677.

Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics. 74:175–195.

Li H. 2011. Improving SNP discovery by base alignment quality. Bioinformatics. 27:1157–1158.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 1303.3997:1–3.

Li H, Handsaker B, Wysoker A, Fennell TJ, Ruan J, Homer N, Marth GT, Abecasis GR, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25:2078–2079.

Li Q, Sun R, Huang C, Wang Z, Liu X, Hou J, Liu J, Cai L, Li N, Zhang S, Wang Y. 2001. Cold adaptive thermogenesis in small mammals from different geographical zones of China. Comp Biochem Physiol Part A Mol Integr Physiol. 129:949–61.

Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE. 2009. On the origin and spread of an adaptive allele in deer mice. Science (80-). 325:1095–1098.

Lischer HEL, Excoffier L, Heckel G. 2014. Ignoring heterozygous sites biases phylogenomic estimates of divergence times: Implications for the evolutionary history of Microtus voles. Mol Biol Evol. 31:817–831.

Lotterhos KE, Whitlock MC. 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol Ecol. 23:2178–2192.

Lotterhos KE, Whitlock MC. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol. 24:1031–1046.

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A. 2017. Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour. 17:142–152.

Lui MA, Mahalingam S, Patel P, Connaty AD, Ivy CM, Cheviron ZA, Storz JF, McClelland GB, Scott GR. 2015. High-altitude ancestry and hypoxia acclimation have distinct effects on exercise capacity and muscle phenotype in deer mice. Am J Physiol - Regul Integr Comp Physiol. 308:R779–R791.

Luu K, Bazin É, Blum MGB. 2017. Pcadapt: An R package to perform genome scans for selection based on principal component analysis. Mol Ecol Resour. 17:67–77.

Lv F-H, Agha S, Kantanen J, Colli L, Stucki S, Kijas JW, Joost S, Li M-H, Marsan PA. 2014. Adaptations to climate-mediated selective pressures in sheep. Mol Biol Evol.

Maas DL, Prost S, Bi K, Smith LL, Armstrong EE, Aji LP, Toha AHA, Gillespie RG, Becking LE. 2018. Rapid divergence of mussel populations despite incomplete barriers to dispersal. Mol Ecol. 27:1556–1571.

Mackay TFC. 2001. The genetic architecture of quantitative traits. Annu Rev Genet. 35:303–339.

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, others. 2012. The *Drosophila melanogaster* genetic reference panel. Nature. 482:173.

Mahalingam S, McClelland GB, Scott GR. 2017. Evolved changes in the intracellular distribution and physiology of muscle mitochondria in high-altitude native deer mice. J Physiol. 595:4785–4801.

Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE. 2010. Convergence in pigmentation at multiple levels: Mutations, genes and function. Philos Trans R Soc B Biol Sci. 365:2439–2450.

Martin A, Orgogozo V. 2013. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution (N Y). 67:1235–1250.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBO J. 17:10.

Martínková N, Barnett R, Cucchi T, Struchen R, Pascal Marine, Pascal Michel, Fischer MC, Higham T, Brace S, Ho SYW, Quere JP, O'Higgins P, Excoffier L, Heckel G, Rus Hoelzel A, Dobney KM, Searle JB. 2013. Divergent evolutionary processes associated with colonization of offshore islands. Mol Ecol. 22:5205–5220.

McClelland GB, Scott GR. 2018. Evolved mechanisms of aerobic performance and hypoxia resistance in high-altitude natives. Annu Rev Physiol. 81:561–583.

McIntyre KM, Setzkorn C, Hepworth PJ, Morand S, Morse AP, Baylis M. 2017. Systematic assessment of the climate sensitivity of important human and domestic animals pathogens in Europe. Sci Rep. 7:7134.

McKenna A, Hanna M, Banks E, Sivachenko AY, Cibulskis K, Kernytsky AM, Garimella K V, Altshuler D, Gabriel S, Daly MJ, DePristo MA. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

McKinney GJ, Larson WA, Seeb LW, Seeb JE. 2017. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al . (2016). Mol Ecol Resour. 17:356–361.

Merilä J. 2012. Evolution in response to climate change: In pursuit of the missing evidence. BioEssays. 34:811–818.

Merilä J, Hendry AP. 2014. Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. Evol Appl. 7:1–14.

Micheletti SJ, Matala AR, Matala AP, Narum SR. 2018. Landscape features along migratory routes influence adaptive genomic variation in anadromous steelhead (*Oncorhynchus mykiss*). Mol Ecol. 27:128–145.

Migalska M, Sebastian A, Konczal M, Kotlík P, Radwan J. 2017. *De novo* transcriptome assembly facilitates characterisation of fast-evolving gene families, MHC class i in the bank vole (*Myodes glareolus*). Heredity (Edinb). 118:348–357.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gatew Comput Environ Work. New Orleans (LA): IEEE; p. 1–8.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17:240–248.

Mills LS, Allendorf FW. 1996. The one-migrant-per-generation rule in conservation and management. Conserv Biol. 10:1509–1518.

De Mita S, Thuillet A-C, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y. 2013. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol Ecol. 22:1383–1399.

Mitton JB, Linhart YB, Hamrick JL, Beckman JS. 1977. Observations on the genetic structure and mating system of ponderosa pine in the Colorado front range. Theor Appl Genet. 51:5–13.

Monge C, Leon-Velarde F. 1991. Physiological adaptation to high altitude: Oxygen transport in mammals and birds. Physiol Rev. 71:1135–1172.

Moreno-Navarrete JM, Ortega F, Gómez-Serrano M, García-Santos E, Ricart W, Tinahones F, Mingrone G, Peral B, Fernández-Real JM. 2013. The MRC1/CD68 ratio is positively associated with adipose tissue lipogenesis and with muscle mitochondrial gene expression in humans.Müller M, editor. PLoS One. 8:e70810.

Müller M, Seifert S, Finkeldey R. 2017. Comparison and confirmation of SNP-bud burst associations in European beech populations in Germany. Tree Genet Genomes. 13:59.

Mzoughi S, Zhang J, Hequet D, Teo SX, Fang H, Xing QR, Bezzi M, Seah MKY, Ong SLM, Shin EM, et al. 2017. PRDM15 safeguards naive pluripotency by transcriptionally regulating WNT and MAPK-ERK signaling. Nat Genet. 49:1354–1363.

Nachman MW, Hoekstra HE, D'Agostino SL. 2003. The genetic basis of adaptive melanism in pocket mice. Proc Natl Acad Sci U S A. 100:5268–5273.

Nadeau S, Meirmans PG, Aitken SN, Ritland K, Isabel N. 2016. The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. Ecol Evol. 6:8649–8664.

Nakamura A, Ebina-Shibuya R, Itoh-Nakadai A, Muto A, Shima H, Saigusa D, Aoki J, Ebina M, Nukiwa T, Igarashi K. 2013. Transcription repressor Bach2 is required for pulmonary surfactant homeostasis and alveolar macrophage function. J Exp Med. 210:2191–2204.

Natarajan C, Jendroszek A, Kumar A, Weber RE, Tame JRH, Fago A, Storz JF. 2018. Molecular basis of hemoglobin adaptation in the high-flying bar-headed goose.Zhang J, editor. PLoS Genet. 14:e1007331.

Natarajan C, Projecto-Garcia J, Moriyama H, Weber RE, Muñoz-Fuentes V, Green AJ, Kopuchian C, Tubaro PL, Alza L, Bulgarella M, et al. 2015. Convergent evolution of hemoglobin function in high-altitude Andean waterfowl involves limited parallelism at the molecular sequence level. PLoS Genet. 11.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A. 76:5269–73.

Nei M, Maruyama T, Chakraborty R. 1975. The bottleneck effect and genetic variability in populations. Evolution (N Y). 29:1.

Nevado B, Ramos-Onsins SE, Perez-Enciso M. 2014. Resequencing studies of nonmodel organisms using closely related reference genomes: Optimal experimental designs and bioinformatics approaches for population genomics. Mol Ecol. 23:1764–1779.

Nielsen R, Korneliussen TS, Albrechtsen A, Li Y, Wang J. 2012. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data.Awadalla P, editor. PLoS One. 7:e37558.

Nilsson JF, Nilsson JÅ. 2016. Fluctuating selection on basal metabolic rate. Ecol Evol. 6:1197–1202.

Norén K, Angerbjörn A. 2014. Genetic perspectives on northern population cycles: Bridging the gap between theory and empirical studies. Biol Rev. 89:493–510.

Nowak R. 1999. Walker’s mammals of the world. Baltimore (MD): Johns Hopkins University Press.

Olsson K, Ågren J. 2002. Latitudinal population differentiation in phenology, life history and flower morphology in the perennial herb *Lythrum salicaria*. J Evol Biol. 15:983–996.

Oue K, Zhang J, Harada-Hada K, Asano S, Yamawaki Y, Hayashiuchi M, Furusho H, Takata T, Irifune M, Hirata M, Kanematsu T. 2016. Phospholipase C-related catalytically inactive protein Is a new modulator of thermogenesis promoted by β-adrenergic receptors in brown adipocytes. J Biol Chem. 291:4185–4196.

Parmesan C, Yohe G. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature. 421:37–42.

Pauls SU, Nowak C, Bálint M, Pfenninger M. 2013. The impact of global climate change on genetic diversity within populations and species. Mol Ecol. 22:925–946.

Pedreschi D, García-Rodríguez O, Yannic G, Cantarello E, Diaz A, Golicher D, Korstjens AH, Heckel G, Searle JB, Gillingham P, et al. 2019. Challenging the European southern refugium hypothesis: Species-specific structures versus general patterns of genetic diversity and differentiation among small mammals. Glob Ecol Biogeogr. 28:262–274.

Peek CB, Levine DC, Cedernaes J, Taguchi A, Kobayashi Y, Tsai SJ, Bonar NA, McNulty MR, Ramsey KM, Bass J. 2017. Circadian clock interaction with HIF1α mediates oxygenic metabolism and anaerobic glycolysis in skeletal muscle. Cell Metab. 25:86–92.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One. 7:e37135.

Petousi N, Croft QPP, Cavalleri GL, Cheng H-Y, Formenti F, Ishida K, Lunn D, McCormack M, Shianna K V., Talbot NP, et al. 2014. Tibetans living at sea level have a hyporesponsive hypoxia-inducible factor system and blunted physiological responses to hypoxia. J Appl Physiol. 116:893–904.

Pevny L, Simon MC, Robertson EJ, Klein WH, Tsai S-FF, D'Agati V, Orkin SH, Costantini F. 1991. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature. 349:257–260.

Phifer-Rixey M, Bi K, Ferris KG, Sheehan MJ, Lin D, Mack KL, Keeble SM, Suzuki TA, Good JM, Nachman MW. 2018. The genomic basis of environmental adaptation in house mice.Payseur BA, editor. PLoS Genet. 14:e1007672.

Poncet BN, Herrmann D, Gugerli F, Taberlet P, Holderegger R, Gielly L, Rioux D, Thuiller W, Aubert S, Manel S. 2010. Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. Mol Ecol. 19:2896–2907.

Prates I, Penna A, Rodrigues MT, Carnaval AC. 2018. Local adaptation in mainland anole lizards: Integrating population history and genome–environment associations. Ecol Evol. 8:11932–11944.

Pritchard JK, Di Rienzo A. 2010. Adaptation – not by sweeps alone. Nat Rev Genet. 11:665–667.

Pritchard VL, Mäkinen H, Vähä JP, Erkinaro J, Orell P, Primmer CR. 2018. Genomic signatures of fine-scale local selection in Atlantic salmon suggest involvement of sexual maturation, energy homeostasis and immune defence-related genes. Mol Ecol. 27:2560–2575.

Pucek Z, Jedrzejewski W, Jedrzejewska B, Pucek M. 1993. Rodent population dynamics in a primeval deciduous forest (Białowieża National Park) in relation to weather, seed crop, and predation. Acta Theriol (Warsz). 38:199–232.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81:559–575.

Puritz JB, Hollenbeck CM, Gold JR. 2014. dDocent : A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ. 2:e431.

Puritz JB, Matz M V., Toonen RJ, Weber JN, Bolnick DI, Bird CE. 2014. Demystifying the RAD fad. Mol Ecol. 23:5937–5942.

R Core Team. 2018. R: A language and environment for statistical computing.

Ratkiewicz M, Borkowska A. 2006. Genetic structure is influenced by environmental barriers: empirical evidence from the common vole *Microtus arvalis* populations. Acta Theriol (Warsz). 51:337–344.

Réale D, Berteaux D, McAdam AG, Boutin S. 2003. Lifetime selection on heritable life-history traits in a natural population of red squirrels. Evolution (N Y). 57:2416–2423.

Redeker S, Andersen LW, Pertoldi C, Madsen AB, Jensen TS, Jørgensen JM. 2006. Genetic structure, habitat fragmentation and bottlenecks in Danish bank voles (*Clethrionomys glareolus*). Mamm Biol - Zeitschrift für Säugetierkd. 71:144–158.

Rellstab C, Fischer MC, Zoller S, Graf R, Tedder A, Shimizu KK, Widmer A, Holderegger R, Gugerli F. 2017. Local adaptation (mostly) remains local: Reassessing environmental associations of climate-related candidate SNPs in *Arabidopsis halleri*. Heredity (Edinb). 118:193–201.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. 2015. A practical guide to environmental association analysis in landscape genomics. Mol Ecol. 24:4348–4370.

Reynolds AW, Mata-Míguez J, Miró-Herrans A, Briggs-Cloud M, Sylestine A, Barajas-Olmos F, Garcia-Ortiz H, Rzhetskaya M, Orozco L, Raff JA, et al. 2019. Comparing signals of natural selection between three Indigenous North American populations. Proc Natl Acad Sci U S A. 116:9312–9317.

Rezende EL, Bozinovic F, Garland T, Garland Jr T. 2004. Climatic adaptation and the evolution of basal and maximum rates of metabolism in rodents. Evolution (N Y). 58:1361–1374.

Richardson JL, Urban MC, Bolnick DI, Skelly DK. 2014. Microgeographic adaptation and the spatial scale of evolution. Trends Ecol Evol. 29:165–176.

Rikalainen K, Aspi J, Galarza JA, Koskela E, Mappes T. 2012. Maintenance of genetic diversity in cyclic populations-a longitudinal analysis in *Myodes glareolus*. Ecol Evol. 2:1491–1502.

Rodríguez A, Rusciano T, Hamilton R, Holmes L, Jordan D, Wollenberg Valero KC. 2017. Genomic and phenotypic signatures of climate adaptation in an *Anolis* lizard. Ecol Evol. 7:6390–6403.

Rohfritsch A, Galan M, Gautier M, Gharbi K, Olsson GE, Gschloessl B, Zeimes C, Vanwambeke SO, Vitalis R, Charbonnel N. 2018. Preliminary insights into the genetics of bank vole tolerance to Puumala hantavirus in Sweden. Ecol Evol. 8:11273–11292.

Rousset F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics. 145:1219–1228.

Roychoudhuri R, Hirahara K, Mousavi K, Clever D, Klebanoff CA, Bonelli M, Sciumè G, Zare H, Vahedi G, Dema B, et al. 2013. BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis. Nature. 498:506–510.

Ruibin W, Zheng X, Chen J, Zhang X, Yang X, Lin Y. 2018. Micro RNA-1298 opposes the effects of chronic oxidative stress on human trabecular meshwork cells via targeting on EIF4E3. Biomed Pharmacother. 100:349–357.

Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. Nat Rev Genet. 14:807–820.

Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. Annu Rev Ecol Evol Syst. 38:595–619.

Sawatani Y, Miyamoto T, Nagai S, Maruya M, Imai J, Miyamoto K, Fujita N, Ninomiya K, Suzuki T, Iwasaki R, et al. 2008. The role of DC-STAMP in maintenance of immune tolerance through regulation of dendritic cell function. Int Immunol. 20:1259–1268.

Saxenhofer M, Schmidt S, Ulrich RG, Heckel G. 2019. Secondary contact between diverged host lineages entails ecological speciation in a European hantavirus.Kuhn JH, editor. PLoS Biol. 17:e3000142.

Schountz T, Acuña-Retamar M, Feinstein S, Prescott J, Torres-Perez F, Podell B, Peters S, Ye C, Black WC, Hjelle B. 2012. Kinetics of immune responses in deer mice experimentally infected with Sin Nombre virus. J Virol. 86:10015–10027.

Schountz T, Quackenbush S, Rovnak J, Haddock E, Black WC, Feldmann H, Prescott J. 2014. Differential lymphocyte and antibody responses in deer mice infected with Sin Nombre hantavirus or Andes hantavirus. J Virol. 88:8319–8331.

Schoville SD, Bonin A, François O, Lobréaux S, Melodelima C, Manel S. 2012. Adaptive genetic variation on the landscape: Methods and cases. Annu Rev Ecol Evol Syst. 43:23–43.

Schweizer M, Excoffier L, Heckel G. 2007. Fine-scale genetic structure and dispersal in the common vole (*Microtus arvalis*). Mol Ecol. 16:2463–2473.

Schweyen H, Rozenberg A, Leese F. 2014. Detection and removal of PCR duplicates in population genomic ddRAD studies by addition of a degenerate base region (DBR) in sequencing adapters. Biol Bull. 227:146–160.

Scott GR, Elogio TS, Lui MA, Storz JF, Cheviron ZA. 2015. Adaptive modifications of muscle phenotype in high-altitude deer mice are associated with evolved changes in gene regulation. Mol Biol Evol. 32:1962–1976.

Scott GR, Schulte PM, Egginton S, Scott ALM, Richards JG, Milsom WK. 2010. Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. Mol Biol Evol. 28:351–363.

Semenza GL. 2007. Hypoxia-inducible factor 1 (HIF-1) pathway. Sci STKE. 2007:cm8–cm8.

Semenza GL. 2009. Regulation of oxygen homeostasis by hypoxia-inducible factor 1. Physiology. 24:97–106.

Sethi JK, Vidal-Puig AJ. 2007. Thematic review series: Adipocyte Biology. Adipose tissue function and plasticity orchestrate nutritional adaptation. J Lipid Res. 48:1253–1262.

Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JBW. 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference.Gilbert M, editor. Methods Ecol Evol. 8:907–917.

Shanmughapriya S, Rajan S, Hoffman NE, Zhang X, Guo S, Kolesar JE, Hines KJ, Ragheb J, Jog NR, Caricchio R, et al. 2015. Ca2+ signals regulate mitochondrial metabolism by stimulating CREB-mediated expression of the mitochondrial Ca2+ uniporter gene MCU. Sci Signal. 8:ra23.

Shenbrot GI, Krasnov BR. 2005. Atlas of the geographic distribution of the arvicoline rodents of the world (Rodentia, Muridae: Arvicolinae). Pensoft.

Sidwell T, Kallies A. 2016. Bach2 is required for B cell and T cell memory differentiation. Nat Immunol. 17:744–745.

Silver DL, Hou L, Somerville R, Young ME, Apte SS, Pavan WJ. 2008. The secreted metalloprotease ADAMTS20 is required for melanoblast survival.Barsh G, editor. PLoS Genet. 4:e1000003.

Simonson TS. 2015. Altitude adaptation: A glimpse through various lenses. High Alt Med Biol. 16:125–137.

Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, et al. 2010. Genetic evidence for high-altitude adaptation in Tibet. Science (80-). 329:72–75.

Skotte L, Korneliussen TS, Albrechtsen A. 2012. Association testing for next-generation sequencing data using score statistics. Genet Epidemiol. 36:430–437.

Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics. 195:693–702.

Slatkin M. 1973. Gene flow and selection in a cline. Genetics. 75:733–756.

Slatkin M. 1987. Gene flow and the geographic structure of natural populations. Science (80-). 236:787–92.

Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G, et al. 2015. The BioMart community portal: An innovative alternative to large, centralized data repositories. Nucleic Acids Res. 43:W589–W598.

Söhle J, Machuy N, Smailbegovic E, Holtzmann U, Grönniger E, Wenck H, Stäb F, Winnefeld M. 2012. Identification of new genes involved in human adipogenesis and fat storage.Rouet P, editor. PLoS One. 7:e31193.

Solari KA, Ramakrishnan U, Hadly EA. 2018. Gene expression is implicated in the ability of pikas to occupy Himalayan elevational gradient.Óvilo C, editor. PLoS One. 13:e0207936.

Song S, Yao N, Yang M, Liu X, Dong K, Zhao Q, Pu Y, He X, Guan W, Yang N, et al. 2016. Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (*Capra hircus*). BMC Genomics. 17:122.

Song H, Luo J, Luo W, Weng J, Wang Z, Li B, Li D, Liu M. 2008. Inactivation of G-protein-coupled receptor 48 (Gpr48/Lgr4) impairs definitive erythropoiesis at midgestation through down-regulation of the ATF4 signaling pathway. J Biol Chem. 283:36687–36697.

Stacy JE, Jorde PE, Steen H, Ims RA, Purvis A, Jakobsen KS. 1997. Lack of concordance between mtDNA gene flow and population density fluctuations in the bank vole. Mol Ecol. 6:751–759.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.

Stawski C, Koteja P, Sadowska ET. 2017. A shift in the thermoregulatory curve as a result of selection for high activity-related aerobic metabolism. Front Physiol. 8:1070.

Stenseth NC. 1999. Population cycles in voles and lemmings: density dependence and phase dependence in a stochastic world. Oikos. 87:427.

Stillwell RC. 2010. Are latitudinal clines in body size adaptive? Oikos. 119:1387–1390.

Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS, Purugganan MD, Schmitt J. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. Proc Natl Acad Sci U S A. 101:4712–4717.

Stojak J, Borowik T, Górny M, McDevitt AD, Wójcik JM. 2019. Climatic influences on the genetic structure and distribution of the common vole and field vole in Europe. Mammal Res. 64:19–29.

Stojak J, McDevitt AD, Herman JS, Kryštufek B, Uhlikova J, Purger JJ, Lavrenchenko LA, Searle JB, Wojcik JM. 2016. Between the Balkans and the Baltic: Phylogeography of a common vole mitochondrial DNA lineage limited to central Europe.Castiglia R, editor. PLoS One. 11:e0168621.

Storz JF. 2002. Contrasting patterns of divergence in quantitative traits and neutral DNA markers: Analysis of clinal variation. Mol Ecol. 11:2537–2551.

Storz JF. 2005. Invited review: Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol Ecol. 14:671–688.

Storz JF. 2016. Hemoglobin-oxygen affinity in high-altitude vertebrates: Is there evidence for an adaptive trend? J Exp Biol. 219:3190–3203.

Storz JF, Cheviron ZA, McClelland GB, Scott GR. 2019. Evolution of physiological performance capacities and environmental adaptation: insights from high-elevation deer mice (*Peromyscus maniculatus*). J Mammal. 100:910–922.

Storz JF, Natarajan C, Cheviron ZA, Hoffmann FG, Kelly JK. 2012. Altitudinal variation at duplicated β-globin genes in deer mice: effects of selection, recombination, and gene conversion. Genetics. 190:203–216.

Storz JF, Sabatino SJ, Hoffmann FG, Gering EJ, Moriyama H, Ferrand N, Monteiro B, Nachman MW. 2007. The molecular basis of high-altitude adaptation in deer mice. PLoS Genet. 3:448–459.

Storz JF, Scott GR, Cheviron ZA. 2010. Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. J Exp Biol. 213:4125–4136.

Strážnická M, Marková S, Searle JB, Kotlík P. 2018. Playing hide-and-seek in beta-globin genes: Gene conversion transferring a beneficial mutation between differentially expressed gene duplicates. Genes (Basel). 9:492.

Sun Y, Hong J, Chen M, Ke Y, Zhao S, Liu W, Ma Q, Shi J, Zou Y, Ning T, et al. 2015. Ablation of Lgr4 enhances energy adaptation in skeletal muscle via activation of Ampk/Sirt1/Pgc1α pathway. Biochem Biophys Res Commun. 464:396–400.

Sutter NB, Karlins E, Davis S, Quignon P, Parker HG, Mosher DS, Ostrander EA, Bustamante CD, Zhu L, Padhukasahasram B, et al. 2007. A single IGF1 allele is a major determinant of small size in dogs. Science (80-). 316:112–115.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.

Takahata S, Isemura N, Mekada K. 2019. 種名が不明なミズハタネズミ亜科系統の種の同定. Proc Okayama Assoc Lab Anim Sci. 35:12–17.

Tarnowska E, Niedziałkowska M, Gerc J, Korbut Z, Górny M, Jędrzejewska B. 2016. Spatial distribution of the Carpathian and Eastern mtDNA lineages of the bank vole in their contact zone relates to environmental conditions. Biol J Linn Soc. 119:732–744.

Tegelström H. 1987. Transfer of mitochondrial DNA from the northern red-backed vole (*Clethrionomys rutilus*) to the bank vole (*C. glareolus*). J Mol Evol. 24:218–227.

Tersago K, Crespin L, Verhagen R, Leirs H. 2012. Impact of Puumala virus infection on maturation and survival in bank voles: A capture-mark-recapture analysis. J Wildl Dis. 48:148–156.

Tersago K, Verhagen R, Servais A, Heyman P, Ducoffre G, Leirs H. 2009. Hantavirus disease (*Nephropathia epidemica*) in Belgium: Effects of tree seed production and climate. Epidemiol Infect. 137:250–256.

Tewksbury JJ, Huey RB, Deutsch CA. 2008. Putting the heat on tropical animals. Science (80-). 320:1296–1297.

Tiffin P, Ross-Ibarra J. 2014. Advances and limits of using population genetics to understand local adaptation. Trends Ecol Evol. 29:673–680.

Tkadlec E, Stenseth NC. 2001. A new geographical gradient in vole population dynamics. Proc R Soc B Biol Sci. 268:1547–1552.

Tkadlec E, Zejda J. 1998. Density-dependent life histories in female bank voles from fluctuating populations. J Anim Ecol. 67:863–873.

Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol. 21:3907–3930.

Tominaga K, Kondo C, Kagata T, Hishida T, Nishizuka M, Imagawa M. 2004. The novel gene fad158, having a transmembrane domain and leucine-rich repeat, stimulates adipocyte differentiation. J Biol Chem. 279:34840–34848.

Tougard C, Renvoisé E, Petitjean A, Quéré J-P. 2008. New insight into the colonization processes of common voles: Inferences from molecular and fossil evidence. PLoS One. 3:e3532.

Travisano M, Mongold JA, Bennett AF, Lenski RE. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. Science (80-). 267:87–90.

Tucker VA. 1970. Energetic cost of locomotion in animals. Comp Biochem Physiol. 34:841–846.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin S V. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nat Genet. 42:260.

Ujvari B, Casewell NR, Sunagar K, Arbuckle K, Wüster W, Lo N, O'Meally D, Beckmann C, King GF, Deplazes E, et al. 2015. Widespread convergence in toxin resistance by predictable molecular evolution. Proc Natl Acad Sci U S A. 112:11911–11916.

Vargas OM, Ortiz EM, Simpson BB. 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: Diplostephium). New Phytol. 214:1736–1750.

Vasconcelos LHC, Souza ILL, Pinheiro LS, Silva BA. 2016. Ion channels in obesity: Pathophysiology and potential therapeutic targets. Front Pharmacol. 7:58.

Vieira FG, Fumagalli M, Albrechtsen A, Nielsen R. 2013. Estimating inbreeding coefficients from NGS data: Impact on genotype calling and allele frequency estimation. Genome Res. 23:1852–1861.

Viitala J, Hakkarainen H, Ylönen H. 1994. Different dispersal in *Clethrionomys* and *Microtus*. Ann Zool Fennici.:411–415.

Villarino A V, Kanno Y, O'Shea JJ. 2017. Mechanisms and consequences of Jak–STAT signaling in the immune system. Nat Immunol. 18:374–384.

de Villemereuil P, Frichot E, Bazin É, François O, Gaggiotti OE. 2014. Genome scan methods against more complex models: When and how much should we trust them? Mol Ecol. 23:2006–2019.

de Villemereuil P, Gaggiotti OE. 2015. A new FST-based method to uncover local adaptation using environmental variables. Methods Ecol Evol.:n/a-n/a.

Visser ME. 2008. Keeping up with a warming world; assessing the rate of adaptation to climate change. Proc R Soc B Biol Sci. 275:649–659.

Waterhouse MD, Erb LP, Beever EA, Russello MA. 2018. Adaptive population divergence and directional gene flow across steep elevational gradients in a climate-sensitive mammal. Mol Ecol. 27:2512–2528.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 7:256–276.

Weber RE. 2007. High-altitude adaptations in vertebrate hemoglobins. Respir Physiol Neurobiol. 158:132–142.

Wei C, Wang H, Liu G, Zhao F, Kijas JW, Ma Y, Lu J, Zhang L, Cao J, Wu M, et al. 2016. Genome-wide analysis reveals adaptation to high altitudes in Tibetan sheep. Sci Rep. 6:26770.

Weigand H, Weiss M, Cai H, Li Y, Yu L, Zhang C, Leese F. 2018. Fishing in troubled waters: Revealing genomic signatures of local adaptation in response to freshwater pollutants in two macroinvertebrates. Sci Total Environ. 633:875–891.

Wellenreuther M, Hansson B. 2016. Detecting polygenic evolution: Problems, pitfalls, and promises. Trends Genet. 32:155–164.

West JB. 1990. Limiting factors for exercise at extreme altitudes. Clin Physiol. 10:265–272.

White TA, Perkins SE, Heckel G, Searle JB. 2013. Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. Mol Ecol. 22:2971–2985.

Whitlock MC, Lotterhos KE. 2015. Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST. Am Nat. 186:S24–S36.

Wójcik JM, Kawałko A, Marková S, Searle JB, Kotlík P. 2010. Phylogeographic signatures of northward post-glacial colonization from high-latitude refugia: A case study of bank voles using museum specimens. J Zool. 281:249–262.

Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK. 2008. Sequence differentiation in regions identified by a genome scan for local adaptation. Mol Ecol. 17:3123–3135.

Woods KA, Camacho-Hübner C, Savage MO, Clark AJL. 1996. Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor I gene. N Engl J Med. 335:1363–1367.

Xing J, Wuren T, Simonson TS, Watkins WS, Witherspoon DJ, Wu W, Qin G, Huff CD, Jorde LB, Ge RL. 2013. Genomic analysis of natural selection and phenotypic variation in high-altitude Mongolians.Williams SM, editor. PLoS Genet. 9:e1003634.

Xu B, Goulding EH, Zang K, Cepoi D, Cone RD, Jones KR, Tecott LH, Reichardt LF. 2003. Brain-derived neurotrophic factor regulates energy balance downstream of melanocortin-4 receptor. Nat Neurosci. 6:736–742.

Yamawaki Y, Oue K, Shirawachi S, Asano S, Harada K, Kanematsu T. 2017. Phospholipase C-related catalytically inactive protein can regulate obesity, a state of peripheral inflammation. Jpn Dent Sci Rev. 53:18–24.

Yan H, Liang C, Li Z, Liu Z, Miao B, He C, Sheng L. 2015. Impact of precipitation patterns on biomass and species richness of annuals in a dry steppe. PLoS One. 10:e0125300.

Yeaman S. 2015. Local adaptation by alleles of small effect. Am Nat. 186:S74–S89.

Yeaman S, Otto SP. 2011. Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. Evolution (N Y). 65:2123–2129.

Yoccoz NG, Stenseth NC, Henttonen H, Prévot-Julliard AC. 2001. Effects of food addition on the seasonal density-dependent structure of bank vole *Clethrionomys glareolus* populations. J Anim Ecol. 70:713–720.

Zhang FL, Shen GM, Liu XL, Wang F, Zhao YZ, Zhang JW. 2012. Hypoxia-inducible factor 1-mediated human GATA1 induction promotes erythroid differentiation under hypoxic conditions. J Cell Mol Med. 16:1889–1899.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, others. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science (80-). 346:1311–1320.

Zhang YY, Yue J, Che H, Sun HY, Tse HF, Li GR. 2014. BKCa and hEag1 channels regulate cell proliferation and differentiation in human bone marrow-derived mesenchymal stem cells. J Cell Physiol. 229:202–212.

Zhigalskii OA. 2011. Structure of the bank vole (*Myodes glareolus*) population cycles in the core and periphery of its species area. Biol Bull. 38:629–641.

Zhong R, Liu H, Wang H, Li X, He Z, Gangla M, Zhang J, Han D, Liu J. 2015. Adaption to high altitude: An evaluation of the storage quality of suspended red blood cells prepared from the whole blood of Tibetan Plateau migrants.Bönig H, editor. PLoS One. 10:e0144201.

Zuur AF, Ieno EN, Elphick CS. 2010. A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol. 1:3–14.

Appendix A.

Supporting material to Chapter 3: Folkertsma et al., in prep
Population genomic evidence for high-altitude adaptation in common voles

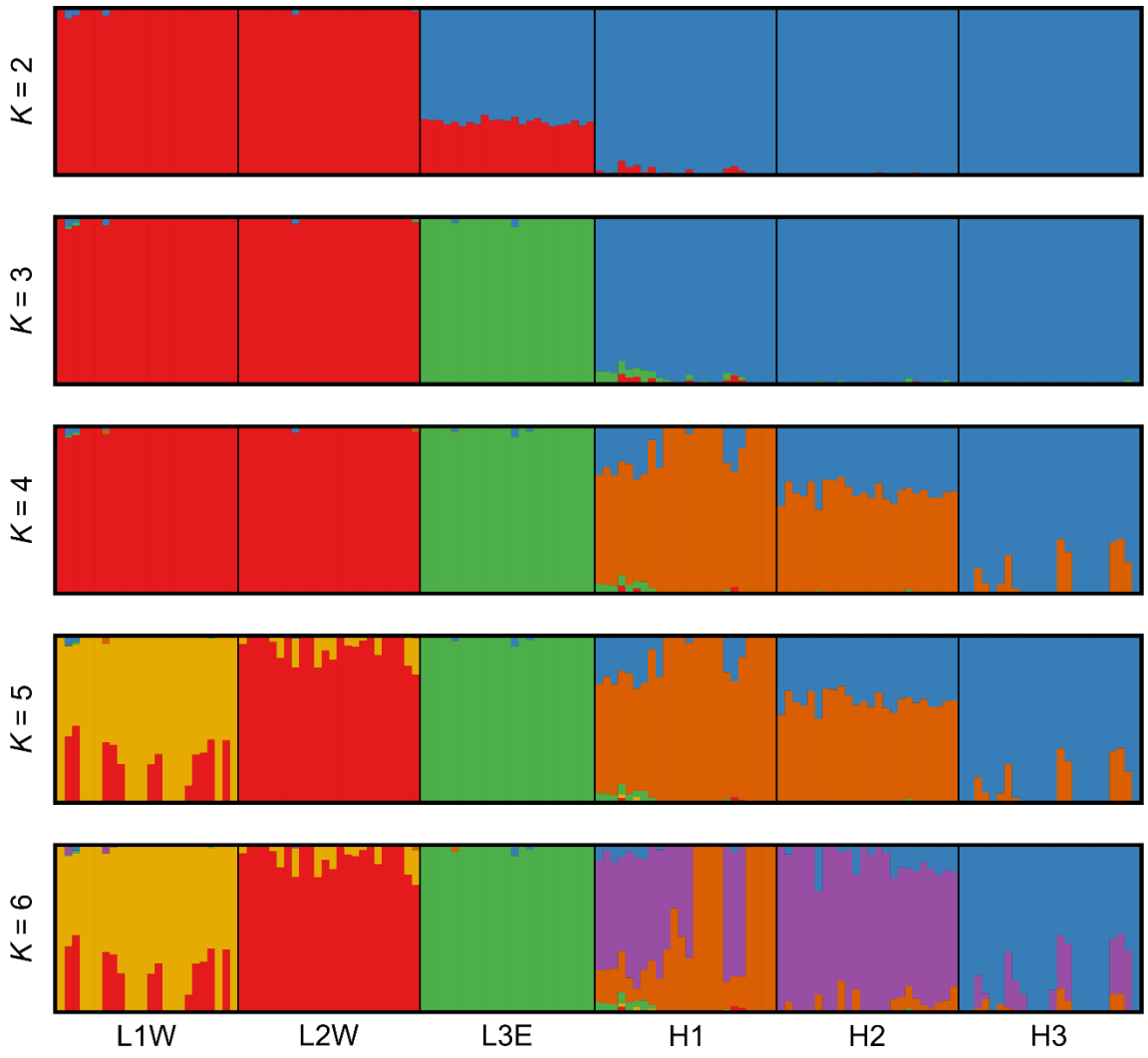


Figure S3.1: Admixture proportions using NgsAdmix based on 19,119 genotyped SNPs, with various numbers of ancestral populations (*K*= 2-6). Each individual is represented by a column with colours corresponding to the proportions of their ancestry component. Vertical black bars separate study sites. Codes for study sites are in Table 3.1.

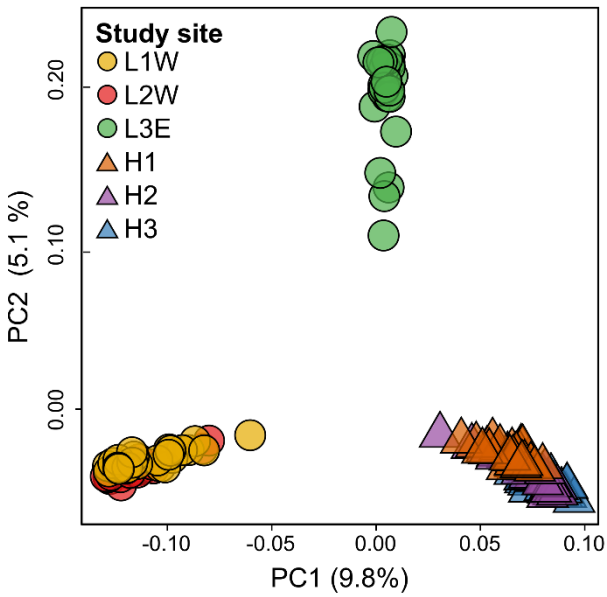


Figure S3.2: Neutral genetic structure of populations as depicted by the first two principal components of a PCA based on pairwise genetic covariance among 143 individuals with 19,119 SNPs 6 *M. arvalis* study sites, together both components explain 14.9% of genetic variation.

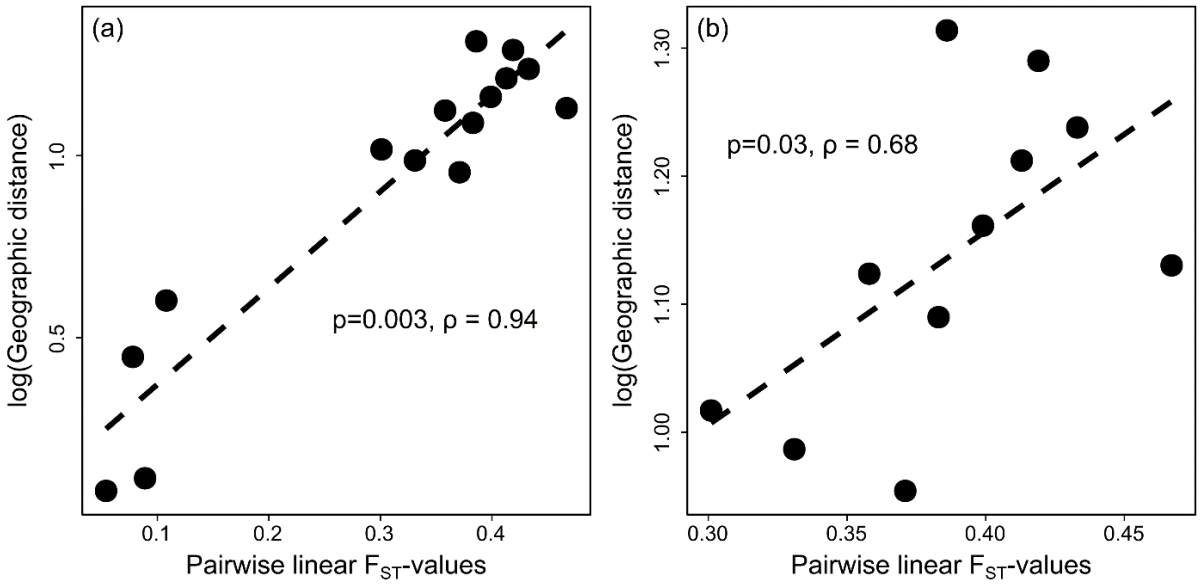


Figure S3.3: Scatterplot showing the results of the Mantel tests between the matrix of genetic distances (Pairwise linear F_{ST}-values) and the matrix of the logarithm of geographic distance, to test for presence of isolation by distance. (a) among all six *M. arvalis* study sites, (b) among geographic distant sites only (by removing pairwise comparison of geographic proximate populations (L1W-L2W; H1-H2; H1-H3; and H2-H3).

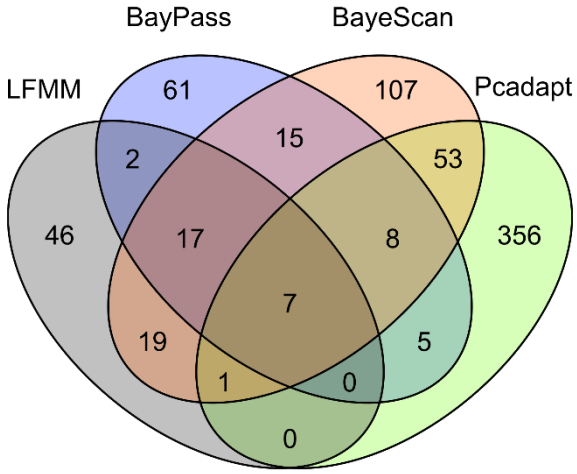


Figure S3.4: Venn diagram comparing outlier SNPs detected by GEA-based methods: LFMM and BayPass, and F_{ST}-based methods: BayeScan and Pcadapt.

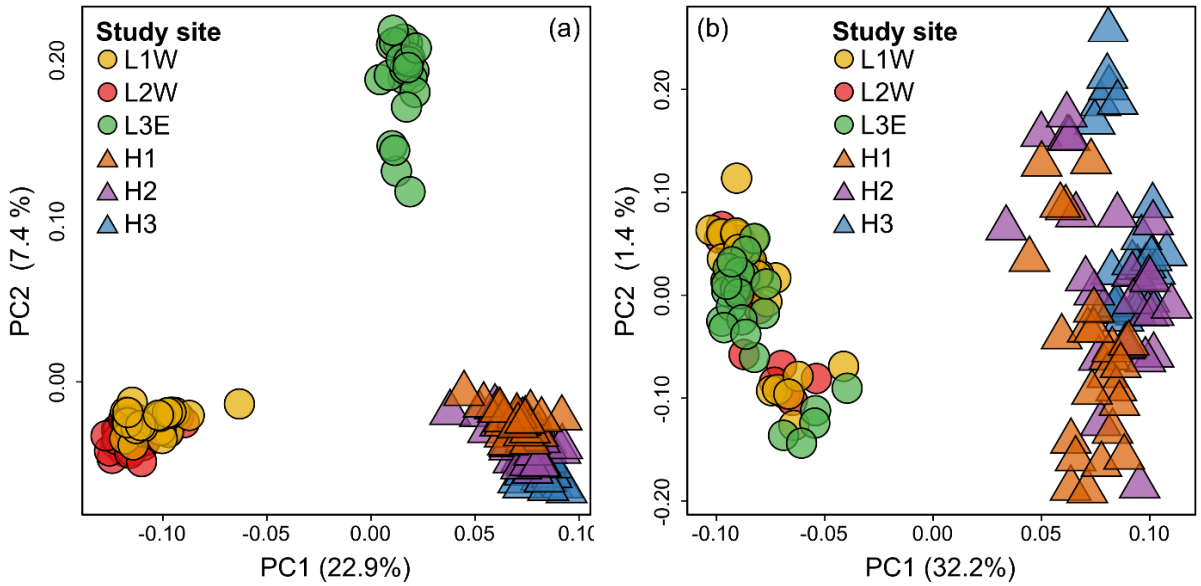
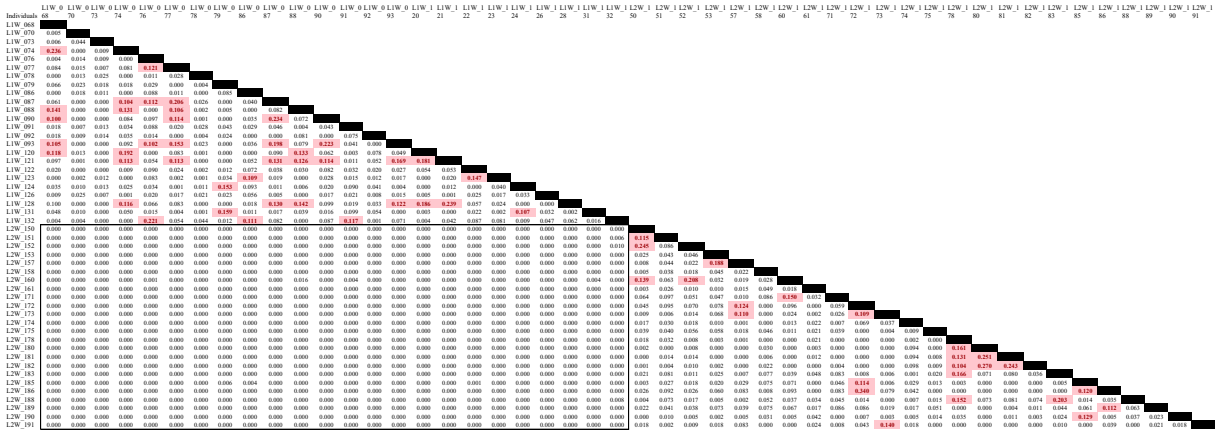


Figure S3.5: Principal component plots of individuals based on pairwise genetic covariance among 143 individuals by using detected outlier SNPs only. (a) PCA plot based on 587 SNPs detected as outliers with F_{ST}-based methods (BayeScan and Pcadapt). (b) PCA plot based on 181 SNPs detected as outliers with GEA-based methods (LFMM and BayPass). The plot created with outliers detected by F_{ST}-based methods is very similar to the neutral PCA plot. While the plot created with outliers detected by GEA-based methods shows a clear separation of individuals in two groups associated with altitude of sampled populations.

Table S3.1: Relatedness estimates for pairs of individuals from the same and proximate study sites. Codes for study sites are in Table 1.



Individuals	L3E_003	L3E_005	L3E_006	L3E_015	L3E_016	L3E_017	L3E_041	L3E_042	L3E_043	L3E_044	L3E_045	L3E_046	L3E_047	L3E_048	L3E_049	L3E_050	L3E_051	L3E_053	L3E_054	L3E_056	L3E_057	L3E_060	L3E_085
L3E_003																							
L3E_005	**0.225**																						
L3E_006	0.000	0.000																					
L3E_015	0.000	0.000	0.000																				
L3E_016	0.004	0.024	0.000	0.016																			
L3E_017	0.000	0.000	0.000	0.000	0.000																		
L3E_041	0.000	0.000	0.000	0.000	0.000	**0.256**																	
L3E_042	0.000	0.000	0.000	0.000	0.000	0.000	0.000																
L3E_043	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000															
L3E_044	0.000	0.000	0.000	0.002	0.000	0.000	0.000	**0.110**	0.000														
L3E_045	0.000	0.000	0.000	0.003	0.011	0.000	0.000	0.000	0.000	0.000													
L3E_046	0.000	0.000	**0.191**	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000												
L3E_047	0.000	0.000	0.024	0.000	0.000	0.000	0.000	0.000	0.000	0.024	0.000	0.053											
L3E_048	0.046	0.026	0.000	0.000	0.000	0.000	0.000	0.034	0.000	0.013	0.000	0.000	0.000										
L3E_049	0.000	0.000	0.000	0.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012									
L3E_050	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.064	0.000	0.083	0.000	0.044	0.000	0.000							
L3E_051	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.007	0.000	0.066	0.000	0.000	0.000	0.000	**0.123**	0.000						
L3E_053	0.000	0.000	0.000	0.041	0.000	0.000	0.000	0.000	0.065	0.001	0.009	0.000	0.000	0.010	0.084	0.000	0.000						
L3E_054	0.000	0.000	0.000	0.017	0.000	0.016	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.046	0.000	0.077	0.000					
L3E_056	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.007	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.093	0.013				
L3E_057	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.096	0.000	0.094	0.000	0.024	0.000	0.000	**0.200**	0.000	0.000	0.000			
L3E_060	0.000	0.000	0.000	0.016	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.034	0.000	0.064	0.000	**0.125**	0.067	0.000		
L3E_085	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.005	**0.118**	0.000	0.000	0.000	0.000	0.057	0.000	0.000	0.000	0.055	0.044	0.067	

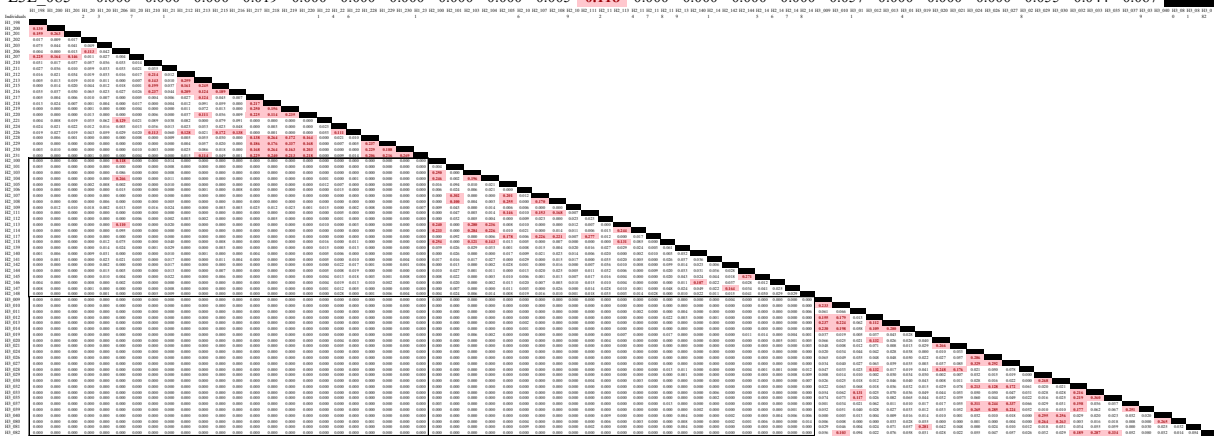


Table S3.2: Above diagonal: between population F_{ST}-values, below diagonal: geographic distance between populations (in kilometers). Codes for study sites are in Table 1.

	L1W	L2W	L3W	H1	H2	H3	Population average
L1W		0.08	0.28	0.26	0.29	0.30	0.24
L2W	1.3		0.30	0.28	0.32	0.29	0.25
L3W	20.6	19.5		0.23	0.25	0.27	0.27
H1	13.3	12.3	10.4		0.05	0.10	0.18
H2	14.5	13.5	9.7	1.2		0.07	0.20
H3	17.3	16.3	9.0	4.0	2.8		0.21

Table S3.3: Significantly over-represented biological processes among 37 candidate genes.

GO term	Biological process	Annotated	Significant	Expected	P-value
GO:0032922	Circadian regulation of gene expression	9	3	0.12	0.00016
GO:0071353	Cellular response to interleukin-4	3	2	0.04	0.00049
GO:0032835	Glomerulus development	14	3	0.18	0.00067
GO:0060142	Regulation of syncytium formation by plasma membrane fusion	4	2	0.05	0.00098
GO:0097028	Dendritic cell differentiation	4	2	0.05	0.00098
GO:0097035	Regulation of membrane lipid distribution	4	2	0.05	0.00098
GO:0042753	Positive regulation of circadian rhythm	4	2	0.05	0.00098
GO:0045599	Negative regulation of fat cell differentiation	6	2	0.08	0.00241
GO:0090263	Positive regulation of canonical Wnt signaling pathway	22	3	0.29	0.00263
GO:0009968	Negative regulation of signal transduction	175	9	2.3	0.00278
GO:0090184	Positive regulation of kidney development	7	2	0.09	0.00335
GO:0043154	Negative regulation of cysteine-type endopeptidase activity involved in apoptotic process	7	2	0.09	0.00335
GO:0045893	Positive regulation of transcription, DNA-templated	206	8	2.7	0.00365
GO:1903845	Negative regulation of cellular response to transforming growth factor beta stimulus	8	2	0.11	0.00443
GO:2001240	Negative regulation of extrinsic apoptotic signaling pathway in absence of ligand	9	2	0.12	0.00565
GO:0061005	Cell differentiation involved in kidney development	10	2	0.13	0.007
GO:0031018	Endocrine pancreas development	10	2	0.13	0.007
GO:0031399	Regulation of protein modification process	236	8	3.1	0.00846

Appendix B.

Supporting material to Chapter 4: Folkertsma et al., in prep
Genomic signatures of climate adaptation in bank voles

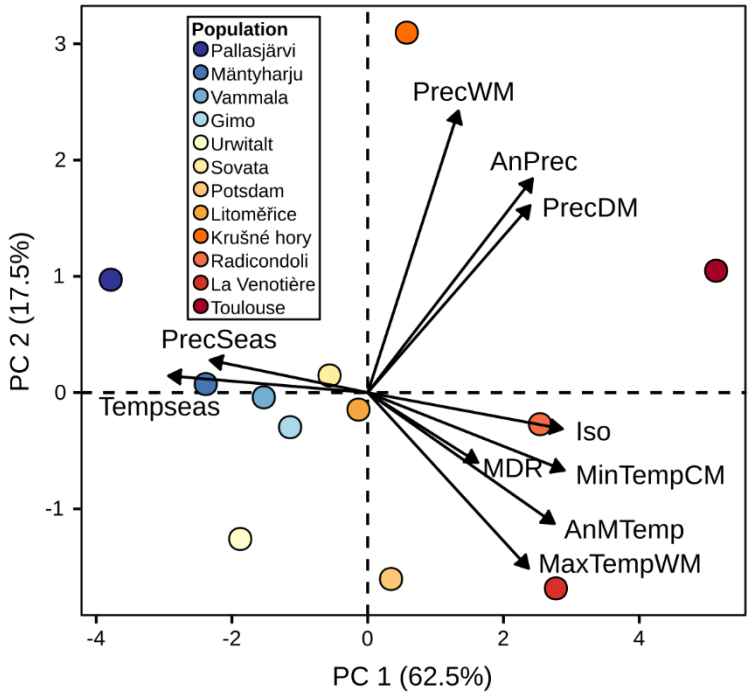


Figure S4.1: Biplot of the principal component analysis depicting relationships between twelve sampled *M. glareolus* populations (colored circles) and 10 climate variables (depicted by arrows). The proportion of total variance explained by each axis is indicated in percent. Environmental identifiers can be found in supplemental table S4.1.

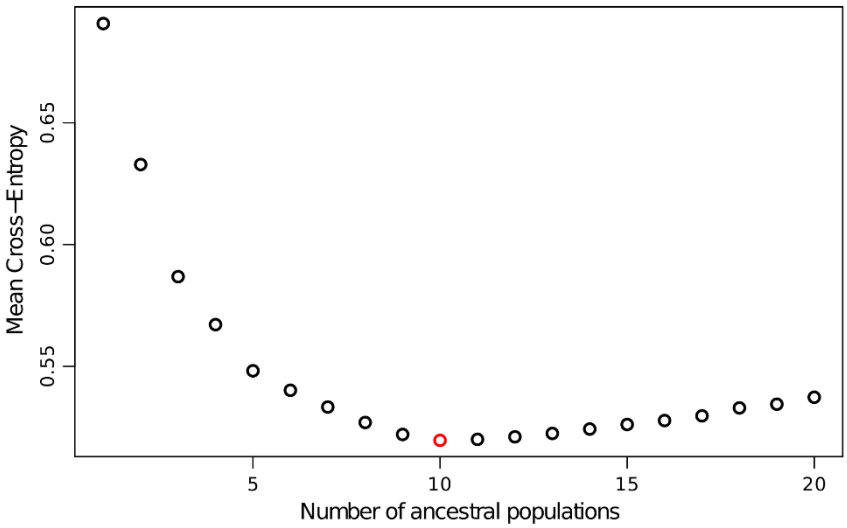


Figure S4.2: Cross-entropy (CE) scores for sNMF runs with the numbers of clusters ranging from 1 to 20. The lowest CE score (*K*=10) is marked in red.

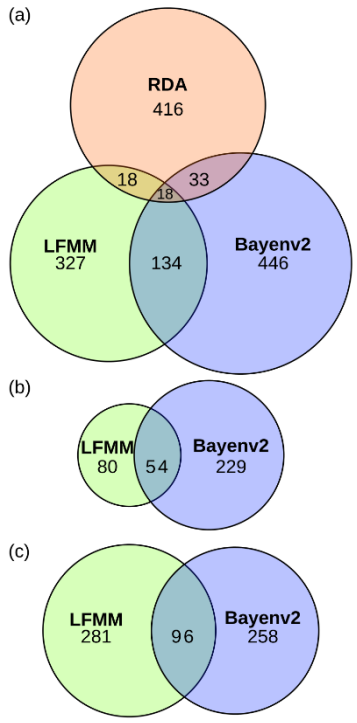


Figure S4.3: Venn diagrams of outlier loci, size of circles represents the number of outlier loci detected. (a) Comparison of overlap between outliers detected by LFMM, Bayenv2 and RDA. Here we used outliers detected using PC1 and PC2 for LFMM and Bayenv2, outliers that were detected by both PCs are included only once. (b) Comparison of overlap between outliers detected by LFMM and Bayenv using PC1. (c) Comparison of overlap between outliers detected by LFMM and Bayenv using PC2.

Table S4.1: Loadings of each climate variable from the WorldClim database to each principal component. For sake of clarity loadings |<0.3| are not presented. Only component 1 and 2 are retained for analysis.

Variable	Description	PC1	PC2	PC3	PC4
AnMTemp	Annual Mean Temperature	0.889	-0.364		
MDR	Mean Diurnal Range (Mean of monthly (max temp - min temp))	0.525		-0.744	0.363
Iso	Isothermality (BIO2/BIO7) (* 100)	0.927			
TempSeas	Temperature Seasonality (standard deviation *100)	-0.948			
MaxTempWM	Max Temperature of Warmest Month	0.766	-0.489		
MinTempCM	Min Temperature of Coldest Month	0.936			
AnPrec	Annual Precipitation	0.785	0.595		
PrecWM	Precipitation of Wettest Month	0.431	0.783	-0.371	
PrecDM	Precipitation of Driest Month	0.774	0.520		
PrecSeas	Precipitation Seasonality (Coefficient of Variation)	-0.751		-0.561	
	Total variance explained per component	62.5%	17.1%	12.7%	5.3%
	Cumulative variance explained	62.5%	79.6%	92.3%	97.6%

Table S4.2: Mapping results of mapped reads onto *M. glareolus* reference genome (GCA_001305785.1) and our improved genome from the Cross-species scaffolding pipeline. We mapped a subset of reads from one sample of each population for reference, resulting in an average increase of properly paired reads of 12.5%.

		M. glareolus reference genome (GCA_001305785.1)	Improved reference genome	
Sample	Population	Number of properly paired reads mapped	Number of properly paired reads mapped	Percentage increase
Pal_03B	Pallasjärvi	1,570,842	1,797,278	12.6
Man_19	Mäntyharju	689,950	799,570	13.7
Vam_581	Vammala	355,088	407,906	12.9
Gim_127	Gimo	844,552	972,112	13.1
Urw_1152	Urwitalt	1,692,884	1,957,572	13.5
Sov_02	Sovata	1,409,272	1,535,772	8.2
Pot_02	Potsdam	1,156,232	1,319,496	12.4
Lit_1914	Litoměřice	514,202	589,764	12.8
Kru_2121	Krušné hory	682,140	785,402	13.1
Rad_01	Radicondoli	344,434	391,868	12.1
Ven_381	La Venotière	607,554	687,584	11.6
Tou_001	Toulouse	858,114	991,050	13.4

Table S4.3: Above diagonal: between population F_{ST}-values, below diagonal: geographic distance between populations (in kilometers). Population IDs can be found in table 4.1.

Distance\F_{ST}	Pal	Man	Vam	Gim	Urw	Sov	Pot	Lit	Kru	Rad	Ven	Tou	Population average F_{ST}
Pal		0.12	0.14	0.39	0.21	0.17	0.26	0.21	0.21	0.36	0.34	0.49	0.27
Man	740		0.04	0.39	0.21	0.17	0.26	0.21	0.20	0.36	0.34	0.48	0.25
Vam	742	216		0.40	0.22	0.18	0.28	0.22	0.22	0.37	0.36	0.50	0.27
Gim	884	499	285		0.33	0.25	0.30	0.25	0.25	0.43	0.40	0.55	0.36
Urw	1,589	910	847	788		0.10	0.18	0.13	0.13	0.33	0.29	0.44	0.23
Sov	2,382	1,657	1,649	1,622	835		0.12	0.07	0.07	0.25	0.21	0.37	0.18
Pot	1,834	1,307	1,158	951	596	1,085		0.10	0.09	0.32	0.29	0.45	0.24
Lit	2,025	1,455	1,325	1,142	640	931	220		0.03	0.29	0.23	0.41	0.2
Kru	2,016	1,455	1,321	1,132	652	962	199	31		0.29	0.23	0.40	0.19
Rad	2,861	2,281	2,161	1,978	1,405	1,166	1,031	838	846		0.29	0.36	0.33
Ven	2,585	2,202	2,020	1,760	1,548	1,767	959	939	917	882		0.22	0.29
Tou	3,081	2,660	2,489	2,240	1,923	1,946	1,368	1,284	1,271	833	507		0.43
Population average distance	1,885	1,398	1,292	1,207	1,067	1,455	973	985	982	1,480	1,462	1,782	

Table S4.4: Summary of candidate genes identified by at least two genome scan methods. Locus refers to the location on the *M. ochgrogaster* reference genome (Chromosome_base-position). Method of outlier detection refers to the genome scan method that identified the SNP as an outlier locus.

Locus	SNP	Gene	GeneID	Method of outlier detection
NC_022009.1_114127352	13905	LOC101979993 - IQCJ-SCHIP1 readthrough transcript protein	101979993	Lfmm_PC2, Bayenv_PC2
NC_022009.1_46105427	8095	Dync1h1 - Dynein cytoplasmic 1 heavy chain 1	101984788	Lfmm_PC1, Bayenv_PC1
NC_022010.1_6202483	3306	Ccdc60 – Coiled-coil domain containing 60	101981129	Lfmm_PC2, Bayenv_PC2
NC_022011.1_19197181	1228	Zfhx3 – Zinc finger homeobox 3	102001415	Lfmm_PC1, Lfmm_PC2, Bayenv_PC1
NC_022011.1_34776034	6118	Pard3 – Par-3 family cell polarity regulator	101989144	Lfmm_PC1, Bayenv_PC1
NC_022011.1_3920185	21260	Sntb2 – Syntrophin beta 2	101992453	Lfmm_PC2, Bayenv_PC2
NC_022011.1_56624398	830	Lrp1b - LDL receptor related protein 1B	101997699	Lfmm_PC2, Bayenv_PC2
NC_022012.1_36407433	10718	Grik4 – Glutamate ionotropic receptor kainate type subunit 4	101999225	Lfmm_PC2, Bayenv_PC2
NC_022012.1_47630215	16807	Exph5 – Exophilin 5	101978966	Lfmm_PC1, Bayenv_PC1, RDA
NC_022012.1_86433265	2658	Glb1 – Galactosidase beta 1	101995815	Lfmm_PC2, Bayenv_PC2
NC_022013.1_34093736	1113	Cdc73 – Cell division cycle 73	102001708	Bayenv_PC1, RDA
NC_022013.1_54264340	11298	Nrg3 – Neuregulin 3	101991068	Lfmm_PC1, Bayenv_PC1
NC_022013.1_58988823	15530	Ptpn20 – Protein tyrosine phosphatase non-receptor type 20	102001893	Bayenv_PC1, RDA
NC_022013.1_70231999	53	Dusp10 – Dual specificity phosphatase 10	101991434	Bayenv_PC1, RDA
NC_022013.1_73321014	625	Gpatch2 - G-patch domain containing 2	101998109	Lfmm_PC1, Bayenv_PC1
NC_022013.1_77189997	3823	Batf3 – Basic leucine zipper ATF-like transcription factor 3	101980506	Lfmm_PC1, Bayenv_PC1, Bayenv_PC2, RDA
NC_022013.1_78476899	1013	Kcnh1 – Potassium voltage-gated channel subfamily H member 1	101985787	Lfmm_PC2, Bayenv_PC2
NC_022014.1_31095426	5192	Rcvrn – Recoverin	101994553	Lfmm_PC2, Bayenv_PC2
NC_022014.1_41285562	8997	Fam114a2 – Family with sequence similarity 114 member A2	101997830	Lfmm_PC1, Bayenv_PC1
NC_022014.1_61163428	21680	Usp32 – Ubiquitin specific peptidase 32	101992481	Lfmm_PC1, Bayenv_PC1
NC_022015.1_15886054	10195	LOC102001996 – Acyl-coenzyme A synthetase ACSM4, mitochondrial	102001996	Lfmm_PC2, Bayenv_PC2
NC_022015.1_2378691	8764	Lyve1 – Lymphatic vessel endothelial hyaluronan receptor 1	101996405	Lfmm_PC2, Bayenv_PC2
NC_022015.1_25578114	5595	Fam53b – Family with sequence similarity 53 member B	101990325	Bayenv_PC1, RDA
NC_022016.1_34744889	10905	Magix - MAGI family member, X-linked	101994189	Bayenv_PC1, RDA
NC_022016.1_65583158	2415	Nfia – Nuclear factor I A	101985327	Lfmm_PC1, Bayenv_PC1
NC_022016.1_6610013	4490	Tnc – Tenascin C	101983607	Lfmm_PC2, Bayenv_PC2
NC_022017.1_15736754	10561	Ift27 – Intraflagellar transport 27	101989188	Lfmm_PC1, Lfmm_PC2, Bayenv_PC1
NC_022017.1_17053779	19263	Adamts20 - ADAM metallopeptidase with thrombospondin type 1 motif 20	101980054	Lfmm_PC2, Bayenv_PC2
NC_022017.1_43070389	18068	Fer1l6 – Fer-1 like family member 6	101996887	Lfmm_PC2, Bayenv_PC2
NC_022018.1_22965832	1502	Camk1d – Calcium/calmodulin dependent protein kinase ID	101997166	Lfmm_PC2, Bayenv_PC2
NC_022018.1_59867563	14768	Ntrk2 – Neurotrophic receptor tyrosine kinase 2	101991185	Lfmm_PC2, Bayenv_PC2, RDA
NC_022019.1_33366136	1761	Dock9 – Dedicator of cytokinesis 9	101983813	Lfmm_PC2, RDA
NC_022020.1_13918386	1319	Dtna – Dystrobrevin alpha	102001187	Lfmm_PC1, Bayenv_PC1, RDA
NC_022020.1_39468798	14030	LOC101982397 – Zinc finger protein 474	101982397	Lfmm_PC2, Bayenv_PC2
NC_022022.1_22815234	9847	Ank2 – Ankyrin 2	101979678	Lfmm_PC2, Bayenv_PC2
NC_022024.1_33246556	5752	Igf1 – Insulin like growth factor 1	101987701	Lfmm_PC2, Bayenv_PC2
NC_022025.1_5022288	5995	Kiaa1324l - KIAA1324 like	101997177	Lfmm_PC2, Bayenv_PC2
NC_022027.1_38831099	13878	Adgrl3 – Adhesion G protein-coupled receptor L3	101986483	Bayenv_PC1, RDA
NC_022027.1_54997482	19286	Lrrc8c – Leucine rich repeat containing 8 VRAC subunit C	101997182	Lfmm_PC2, RDA
NC_022028.1_28663654	13488	St6gal2 - ST6 beta-galactoside alpha-2,6-sialyltransferase 2	101979687	Lfmm_PC2, Bayenv_PC2
NC_022028.1_33292795	9202	Cdh23 – Cadherin related 23	101999180	Lfmm_PC1, Bayenv_PC1
NC_022028.1_35652758	1125	Rufy2 - RUN and FYVE domain containing 2	101987900	Lfmm_PC2, Bayenv_PC2
NC_022028.1_49631929	8437	Dnah8 – Dynein axonemal heavy chain 8	101988652	Lfmm_PC2, Bayenv_PC2
NC_022029.1_13954523	5544	Atad2b - ATPase family AAA domain containing 2B	101991203	Lfmm_PC2, Bayenv_PC2
NC_022030.1_4789630	16537	Slc2a12 – Solute carrier family 2 member 12	101990267	Bayenv_PC1, RDA
NC_022030.1_57114683	9559	Tns1 – Tensin 1	102001111	Lfmm_PC1, Bayenv_PC1
NC_022031.1_25142086	17315	Fut9 – Fucosyltransferase 9	101991953	Lfmm_PC2, Bayenv_PC2
NC_022031.1_30606927	2299	Bach2 - BTB domain and CNC homolog 2	101995098	Lfmm_PC1, Lfmm_PC2, Bayenv_PC2, RDA
NC_022033.1_117651	13141	Myt1 – Myelin transcription factor 1	101989790	Bayenv_PC1, RDA
NC_022033.1_15962378	375	Ptprt – Protein tyrosine phosphatase receptor type T	101998531	Lfmm_PC2, Bayenv_PC1, RDA
NC_022034.1_10692654	12357	Fndc1 – Fibronectin type III domain containing 1	101985076	Lfmm_PC2, Bayenv_PC2
NC_022034.1_34066290	6313	Tbc1d32 - TBC1 domain family member 32	101979138	Bayenv_PC2, RDA
NC_022034.1_40747119	13061	Ptprk – Protein tyrosine phosphatase receptor type K	101989309	Lfmm_PC2, Bayenv_PC2
NC_022036.1_5481645	16269	LOC101998720 – Anionic trypsin-2	101998720	Bayenv_PC1, RDA
NW_004949095.1_788514	7990	C6H1orf21 – Chromosome 6 C1orf21 homolog	101983365	Lfmm_PC2, Bayenv_PC2
NW_004949096.1_35517029	20241	Blvra – Ciliverdin reductase A	101996347	Lfmm_PC2, Bayenv_PC2
NW_004949098.1_1764557	14228	Mrpl19 – Mitochondrial ribosomal protein L19	101984889	Lfmm_PC1, RDA
NW_004949099.1_1813079	16762	Aplf – Aprataxin and PNKP like factor	101983646	Lfmm_PC1, Bayenv_PC1
NW_004949099.1_24984757	14356	Grm7 – Glutamate metabotropic receptor 7	102002042	Lfmm_PC2, Bayenv_PC2
NW_004949099.1_7212850	11906	LOC113458036 – Uncharacterized LOC113458036	113458036	Lfmm_PC1, Bayenv_PC1, RDA
NW_004949102.1_10543126	731	Exoc4 – Exocyst complex component 4	101999828	Lfmm_PC1, Lfmm_PC2, Bayenv_PC2, RDA
NW_004949102.1_1626706	3317	LOC101997986 – Hyaluronidase-5-like	101997986	Bayenv_PC1, RDA
NW_004949102.1_4439720	7323	Grm8 – Glutamate metabotropic receptor 8	102000941	Lfmm_PC1, Bayenv_PC1
NW_004949102.1_6516662	13798	Ahcyl2 – Adenosylhomocysteinase like 2	101989422	Bayenv_PC1, RDA
NW_004949106.1_3445183	19410	Plcl1 – Phospholipase C like 1 (inactive)	101997890	Bayenv_PC1, RDA
NW_004949106.1_7957839	21309	Stat4 – Signal transducer and activator of transcription 4	101994625	Lfmm_PC2, Bayenv_PC2
NW_004949111.1_3852573	15198	LOC101997307 – Uncharacterized LOC101997307	101997307	Lfmm_PC1, RDA
NW_004949117.1_4850355	15033	Dpys – Dihydropyrimidinase	101985105	Bayenv_PC1, RDA
NW_004949128.1_2823627	5956	Atp10b - ATPase phospholipid transporting 10B (putative)	101983489	Lfmm_PC2, Bayenv_PC2
NW_004949130.1_2298948	9797	LOC101979737 – Keratin, type I cytoskeletal 18 pseudogene	101979737	Lfmm_PC2, Bayenv_PC2
NW_004949149.1_1068966	11743	Inpp4b – Inositol polyphosphate-4-phosphatase type II B	101992938	Lfmm_PC1, Lfmm_PC2
NW_004949155.1_1652207	7752	Slc16a2 – Solute carrier family 16 member 2	101982460	Lfmm_PC1, Bayenv_PC1
NW_004949164.1_394094	19532	Elmo1 – Engulfment and cell motility 1	101998663	Bayenv_PC1, RDA
NW_004949242.1_70943	16243	LOC101986078 – Vomeronasal type-1 receptor 4-like	101986078	Lfmm_PC1, Bayenv_PC1

Table S4.5: Percentage of outliers associated with climate variables based on pRDA for different datasets containing: all 1,392 outliers detected, only 485 outliers detected by RDA, and 213 selected outliers detected by at least two methods.

	All 1,392 outliers	485 RDA outliers	213 selected outliers
AnMTemp	27.5%	37.1%	28.6%
TempSeas	14.8%	11.1%	15.0%
MDR	18.1%	20.8%	13.6%
AnPrec	25.4%	15.9%	27.7%
PrecSeas	14.2%	15.1%	15.0%

Table S4.6: Correlation between climate variables from the WorldClim database. Above diagonal: correlation coefficient (spearman's ρ), below diagonal significance level of correlations (NS=non-significant, * < 0.05, ** < 0.01, *** < 0.001. Abbreviations can be found in Appendix B: table S4.1.

	AnM Temp	MDR	Iso	Temp Seas	Max TempWM	Min TempCM	AnPrec	PrecWM	PrecDM	PrecSeas
AnMTemp		0.41	0.82	-0.83	0.93	0.93	0.47	0.04	0.55	-0.60
MDR	NS		0.67	-0.31	0.60	0.29	0.46	0.32	0.27	-0.22
Iso	**	*		-0.86	0.77	0.82	0.73	0.35	0.70	-0.63
TempSeas	**	NS	***		-0.68	-0.94	-0.72	-0.23	-0.84	0.85
MaxTempWM	***	*	**	*		0.81	0.44	0.01	0.44	-0.51
MinTempCM	***	NS	**	***	**		0.61	0.15	0.78	-0.79
AnPrec	NS	NS	**	**	NS	*		0.72	0.80	-0.53
PrecWM	NS	NS	NS	NS	NS	NS	**		0.39	0.02
PrecDM	NS	NS	*	***	NS	**	**	NS		-0.86
PrecSeas	*	NS	*	***	NS	**	NS	NS	***	

Declaration of authorship

I, Remco Folkertsma, hereby declare that the work presented within this thesis titled, "Evolutionary adaptation to climate in microtine mammals", is purely my own unless stated otherwise.

Chapter 1:

"The complete mitochondrial genome of the common vole, *Microtus arvalis* (Rodentia: Arvicolinae)"
This manuscript was published in "Mitochondrial DNA partB" in 2018 under http://doi.org/10.1080/23802359.2018.1457994. I carried out lab work, phylogenetic analysis and wrote the manuscript. Michael Westbury assembled the mitochondrial genome. All authors contributed helpful input to the manuscript.

Chapter 2:

"Population genetic evidence for high-altitude adaptation in common voles"
This manuscript has been submitted to "Molecular Ecology" in December 2019. I conceived the project idea together with Gerald Heckel, Jana Eccard and Michael Hofreiter. I performed field work and collected samples. I carried out the vast majority of lab work. I performed all data analyses. I wrote the majority of the manuscript with valuable input from Gerald Heckel, Jana Eccard, Michael Hofreiter and Johanna Paijmans.

Chapter 3:

"Genomic signatures of climate adaptation in bank voles"
This manuscript is in preparation for submission to the "Journal of Evolutionary Biology". I conceived the project idea together with Jana Eccard and Michael Hofreiter. I performed field work and collected some of the samples. I carried out the vast majority of lab work. I performed all data analyses. I wrote the majority of the manuscript with valuable input from Michael Hofreiter.

Signed:

_____

Date:

_____